The Dissertation Committee for Peck-Sheng Wee

Certifies that this is the approved version of the following dissertation:

Effects of Geographic Distance, Landscape Features and Host Association on Genetic Differentiation of Checkerspot Butterflies.

Committee:

_____
Michael C. Singer, Supervisor

_____
Ulrich G. Mueller, Co-supervisor

_____
Craig R. Linder

_____
Mark Kirkpatrick

_____
Camille Parmesan

Effects of Geographic Distance, Landscape Features and Host Association on Genetic Differentiation of Checkerspot Butterflies.

by

Peck-Sheng Wee, B.Sc., M.Sc.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

In Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2004

Effects of Geographic Distance, Landscape Features and Host Association on Genetic Differentiation of Checkerspot Butterflies.

Publication No. _____

Peck-Sheng Wee, Ph.D.
The University of Texas at Austin, 2004

Supervisor: Michael C. Singer
Co-supervisor: Ulrich G. Mueller

Wright's hypothesis of genetic isolation by distance (IBD) predicts that genetic differentiation increases with geographic distance. The use of different diets by allopatric populations may also contribute significantly to genetic isolation by restricting gene flow among populations using different hosts. The overarching objective is to ask whether and to what extent the degree of isolation between populations is associated with differences between them in host use after taking into account the effects of spatial isolation. AFLP markers were used to assess population genetic structure. In the first study, I ask if there is detectable host associated genetic differentiation between allopatric populations of the marsh fritillary butterfly *Euphydryas aurinia*. Each of the ten sampled populations utilizes one of two hosts, and an additional population utilizes both hosts. Substantial host associated genetic differentiation was found between allopatric populations (F_{CT} = 0.12, $P < 0.001$), but no such association could be found within the dual host population. In the second study, I use a GIS to ask if whether and to what extent the degree of genetic differentiation between *Euphydryas editha* populations is associated with the type of intervening landscape matrix. Eight populations each from Sequoia National Park (SEKI) and Yosemite National Park (YOSE) were sampled. Genetic differentiation within SEKI

and YOSE was found to be small but statistically significant ($\Phi_{ST} = 0.03$, $P < 0.001$). At SEKI, no IBD was detected. In contrast, IBD was detected at YOSE ($r = 0.66$, $P < 0.001$). When landscape features were taken into account, the IBD relationship was strengthened ($r = 0.77$, $P < 0.001$). In the third study, an additional 26 populations were sampled to assess regional scale differentiation. There was substantial genetic differentiation between populations ($\Phi_{ST} = 0.19$, $P < 0.001$) and significant IBD ($r = 0.55$, $P < 0.001$). Host association was also found to be significantly correlated with genetic differentiation ($r = 0.17$, $P < 0.01$). Populations on *Castilleja* exhibited a significant IBD relationship ($r = 0.7$, $P < 0.001$). Populations on *Collinsia*, on the other hand, appear to be recently colonized, and do not exhibit IBD.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.0 BACKGROUND

Wright (1943) was one of the first to suggest a relationship between genetic differentiation and geographical separation. Subsequent work in theoretical population genetics modeled the relative importance of selection, migration, drift, and mutation on the evolution of traits in natural populations. MacArthur and Wilson (1963; 1967) similarly advanced ecological theory by elucidating the role of physical geography on community structure and the dynamics of immigration and extinction. The modern field of metapopulation ecology introduced a new level of sophistication by exploring the roles that habitat size, habitat quality, fragmentation, and connectivity play on colonization and extinction dynamics among locally interacting populations.

In this dissertation, I synthesize concepts from landscape ecology and molecular ecology to investigate the population genetic structure of the butterfly *Euphydryas editha* in the Sierra Nevada. Populations in the Sierra Nevada utilize five different host genera for oviposition. Population structure is variable within this species: those found in Sequoia National Forest and Yosemite National Park typically exist as large metapopulations, while other ecotypes comprise well defined and relatively isolated populations. The sedentary nature and patchy distribution of *E. editha* populations (Ehrlich 1965; Harrison 1989), combined with known spatial variation in their use of hosts (Singer 1994; Singer and Thomas 1996) makes this an ideal system for studies on relationships between population structure, resource adaptation, and genetic differentiation. Different predictions can be made about the evolution and differentiation of *E. editha* populations, depending on the relative importance of variables such as diet and geographical separation.

The principle question I address in this dissertation is: What are the relative effects of local adaptation and geographic separation on the population genetic structure of *E. editha* populations? To that end, this dissertation also includes a study based on *Euphydryas aurinia*, a closely related butterfly species found throughout Europe. For this study, I focused on a group of populations in southern France and north-eastern Spain which are known to utilize only two host genera. The inclusion of these populations serves as a contrast to the *E. editha* populations which utilize five host genera.

2.0 ORGANIZATION

Chapter 2 describes the molecular methods common to all the studies in this dissertation. The molecular marker employed was amplified fragment length polymorphism (AFLP). A frequent problem encountered by researchers dealing with small quantities of insect tissue sample is AFLP data that are difficult to analyze due to low PCR efficiency. I describe modified DNA extraction and AFLP-PCR protocols which produce reliable AFLP profiles. I also describe a custom written program – AFLPal – which reduces the subjectivity involved in the scoring of AFLP markers and substantially reduces the overall AFLP editing time.

Chapter 3 is inspired by the recent resurgence of interest in sympatric speciation in herbivorous insects. A number of studies have found host-associated genetic differentiation between sympatric "host races" (Feder et al. 1993; Feder et al. 1994; Via 1999; Abrahamson et al. 2001). They have also recorded host-associated selection that promotes such differentiation and various classes of pre-mating and post-mating isolation that help maintain it. *E. aurinia*, in contrast to other herbivorous insects that have so far been studied in the speciation context, exhibits little evidence of host-associated isolation. Nevertheless, I found a substantial amount of host-associated genetic differentiation between allopatric groups of populations found on either one of two hosts. This effect was not found at a population which utilizes both hosts.

Chapter 4 details the first known attempt at correlating genetic differentiation with the effects of landscape heterogeneity using information on the relative resistance posed by landscape elements. Isolation by distance (IBD), the reduction of gene flow with increasing geographic separation, can cause genetic differentiation to increase with geographic distance. For organisms with patchy spatial distributions, this effect can be augmented by the influence of the matrix of unsuitable landscape that separates suitable patches. The objective of this study is to investigate whether a Geographical Information System can be used to model the influence of the matrix on effective isolation between *E. editha* populations. Eight populations each from Sequoia National Park / National Forest and Yosemite National Park were sampled and assayed using AFLP markers. ArcGIS was used to perform cost-surface modeling to test three metrics of effective isolation: accumulated costs, topographical distances along least cost paths, and least cost path lengths. These metrics were generated using three ArcGIS cost models that make different assumptions about how landscape features contribute to effective isolation. At Yosemite, it was found that accumulative costs computed with the most basic cost model resulted in an increased correlation between genetic differentiation and spatial distances.

Chapter 5 addresses the combined effects of local adaptation and geographic separation on population differentiation at two different geographical scales. There have been numerous studies investigating IBD to date (reviewed in Peterson and Denno 1998). There have also been studies investigating host-associated genetic differentiation, but to date, there has been no investigation into the relative effects of spatial and ecological barriers on the genetic structure of natural insect populations. Differences among allopatric populations in host utilization may contribute significantly to their genetic isolation. Genetic isolation of populations is also influenced by their geographic separation. In this study, I sampled populations at the regional and local scale, where populations are separated by an average of 259km and 9km respectively. I found substantial genetic differentiation between populations at the regional scale. A smaller degree of differentiation was found at the local scale within Sequoia National Park /

National Forest and Yosemite National Park. Genetic differentiation at the regional scale and at Yosemite was found to be correlated with both geographical separation and differences in host association between populations. At the regional scale, populations utilizing host plant species belonging to the genus *Castilleja* (subgenus *castilleja)* exhibited a strong and significant IBD relationship, in contrast to populations utilizing host plant species belonging to the genus *Collinsia*. These latter populations, which appear to be recently colonized, do not exhibit IBD.

References

Abrahamson, W. G., M. D. Eubanks, et al. (2001). "Gall flies, inquilines, and goldenrods: A model for host-race formation and sympatric speciation." American Zoologist **41**(4): 928-938.

Ehrlich, P. (1965). "The population biology of the butterfly *Euphydryas editha.*" Evolution **19**: 327-336.

Feder, J. L., T. Hunt, et al. (1993). "The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella.*" Entomologia Experimentalis et Applicata **69**: 117-135.

Feder, J. L., S. Opp, et al. (1994). "Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly." Proc Natl Acad Sci USA **91**: 7990-7994.

Harrison, S. (1989). "Long-distance dispersal and colonization in the bay checkerespot butterfly, *Euphydryas editha bayensis*." Ecology **70**: 1236-1243.

MacArthur, R. and E. Wilson (1963). "An equilibrium theory of insular zoogeography." Evolution **17**: 373-387.

MacArthur, R. and E. Wilson (1967). The Theory of Island Biogeography. Princeton, New Jersey., Princeton University Press.

Peterson, M. and R. Denno (1998). "The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects." American Naturalist **152**(3): 428-446.

Singer, M. C. (1994). Behavioral constraints on the evolutionary expansion of insect diet: a case history from checkerspot butterflies. Behavioral Mechanisms in Evolutionary Ecology. L. Real, University of Chicago Press**:** 279-296.

Singer, M. C. and C. D. Thomas (1996). "Evolutionary responses of a butterfly metapopulation to human and climate-caused environmental variation." American Naturalist **148**: S9-S39.

Via, S. (1999). "Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice." Evolution **53**(5): 1446-1457.

Wright, S. (1943). "Isolation by distance." Genetics **28**: 139-156.

Chapter 2: An Efficient AFLP Protocol for Butterflies: from Laboratory to Analysis.

Synopsis: This chapter describes the molecular methods common to the studies described in later chapters. The molecular marker of choice was amplified fragment length polymorphism (AFLP). These markers are generated using a manufacturer supplied polymerase chain reaction (PCR) protocol. A frequent problem encountered by researchers dealing with small quantities of insect tissue sample is AFLP data that are difficult to analyze due to low PCR efficiency. I describe modified DNA extraction and AFLP-PCR protocols which produce reliable AFLP profiles. Once generated, the AFLP marker data needs to be edited before it is used in statistical analyses. The editing of AFLP markers using the manufacturer supplied program, Genotyper v3.6, is tedious, subjective and error prone. This is especially true when dealing with large sample sizes with hundreds of AFLP markers. I describe a program – AFLPal – which when used in association with a set of methods in Genotyper, reduces the subjectivity involved in the scoring of AFLP markers and substantially reduces the overall editing time. An alternate Visual Basic for Applications (VBA) for Excel called AFLPexel is available as an alternative to AFLPal for less demanding AFLP applications.

2.1 INTRODUCTION

Amplified Fragment Length Polymorphism (AFLP) markers are DNA regions flanked by a pair of primers of known sequence. The AFLP protocol uses a pair of primers corresponding to those sequences to generate DNA fragments wherever those sequences occur in the genome. The resultant data are essentially a reproducible profile of the assayed specimen assembled from random parts of its genome. AFLP markers offer a series of advantages over other standard markers such as allozymes, RFLPs, minisatellites, and microsatellites (Jones et al. 1997; Mueller and Wolfenbarger 1999)

and are useful for assessing genetic differences among individuals, populations, and closely related species (Mueller and Wolfenbarger 1999).

Despite its ease of use, the technique can prove challenging when attempting to assay small amounts of DNA. I present a modified DNA extraction process (based on a Qiagen extraction kit) and a modified AFLP protocol (based on the Applied Biosystems protocol) that has worked very reliably when the amount of material available for DNA extraction is small. The modified protocols have been successfully been used on *Euphydryas editha* (*Nymphalidae*) third instar larval heads, butterfly wing clippings measuring no more than 0.5cm on a side, single butterfly legs, and phorid flies (*Pseudacteon curvatus*). Prior to adoption of this new set of protocols, AFLP reactions from phenol-chloroform extracted specimens failed 30 – 40% of the time (personal observation). In contrast, success rates with this new set of protocols are typically over 90%, leading to a substantial savings in time and money.

Even if AFLP reactions are repeatable and produce usable data, editing the raw AFLP data for analysis by population genetic programs is laborious, time consuming, and often subjective. The main challenge in this step is to identify homologous markers in order to score individuals reliably. I present an overview of a custom program, AFLPal, which may be used to accelerate and improve the accuracy of editing raw AFLP data generated by the ABI 3100 (or similar) automated genetic analyzers. AFLPal is designed to overcome certain shortfalls with version 3.6 of the ABI Genotyper program. Later releases of the Genotyper program may have added features that address the shortcomings presented here. This chapter is not meant to be a user manual for either Genotyper v3.6 or AFLPal, but is intended to highlight features of both programs that may help optimize the editing process.

2.2 MODIFIED LABORATORY PROTOCOLS

DNA extraction. Qiagen DNeasy Tissue kits (Catalog No. 69506) were used for DNA extraction. Instead of using Qiagen's tissue lysis buffer, a CTAB extraction buffer (10% 1 M Tris-HCL pH 8.0, 28% 5 M NaCl, 4% 0.5 M EDTA pH 8.0, 2% CTAB) was used. Prior to extraction, 2 µl of β-mercaptoethanol per 1 ml of extraction buffer was added to the buffer. Samples were crushed in a 1.5 ml Eppendorf tube in liquid nitrogen, and then homogenized in 200 µl of the extraction buffer. The Qiagen protocol was followed for the rest of the steps. The DNA was eluted in two 100 µl washes of buffer AE for a final DNA volume of 200 µl.

AFLP PCR reaction. The Applied Biosystems (ABI) AFLP protocol for small plant genomes (www.appliedbiosystems.com; protocol #: 4303146) was used with the following modifications. 30 ng of genomic DNA was used for restriction-ligation instead of the recommended 500 ng. The ABI protocol suggests either overnight benchtop incubation or 37^{o}C incubation for 2 hours using a thermocycler with a heated lid (at 37^{o}C) for the restriction-ligation reaction. The former frequently results in a high proportion of blank AFLP samples. Incubation at 37^{o}C on a thermocycler with a heated lid (at 37^{o}C) for 3 hours results in substantially improved results. Prior to the pre-selective amplification, 120 µl of $TE_{0.1}$ was added to the restriction-ligation product instead of the recommended 189 µl. For the pre-selective amplification, the regular genome pre-selective primer (www.appliedbiosystems.com; product #: 402004) was found to work very well for certain insect genera (*Euphydryas*, *Heliconius*, and *Pseudacteon)*. For this step, 25 cycles of 20 s at 94 °C, 30 s at 56 °C, 2 min at 72 °C were used instead of the recommended 20. The pre-selective product was not diluted for the selective amplification. For the selective PCR, 25 cycles of 20 s at 94 °C, 30 s at 56 °C, 2 min at 72 °C were used instead of the recommended 20. 3 µl of the selective product was mixed with 0.4 µl of GeneScan 500 Rox and 6.6 µl of Hi-Di Formamide. Samples were processed on an automated ABI PRISM 3100 Genetic Analyzer.

2.3 CHALLENGES IN THE AFLP SCORING PROCESS

2.3.1 Overview

The output from the ABI PRISM 3100 Genetic Analyzer (or equivalent) comprises a series of AFLP profiles, one for each assayed sample. When imported into the program Genotyper, each profile appears as a series of peaks (Figure 2.1). Each peak corresponds to a fluorescently labeled AFLP fragment (AFLP marker). Depending on the primer set used, the number of fragments may vary from the low tens to two hundred (personal observation). The size of each fragment, in base pairs (bp), is determined through an interpolation process using fragments of known sizes. The height of each peak in Genotyper corresponds to the amount of PCR product of a particular size.

2.3.2 Challenges in Fragment Categorization

The objectives of the scoring process are to (1) identify homologous fragments, and (2) label these fragments to produce an absence-presence matrix that represents, for each individual specimen, the absence or presence of a given fragment.

Identifying homologous fragments. The process of identifying homologous fragments and assigning them to discrete categories is one of the first steps in scoring AFLP data. Each category corresponds to an AFLP locus. These fragments appear as 'peaks' when viewed in Genotyper (Figure 2.1). Accordingly, homologous fragments are said to belong to the same "peak category" of a given size. One would expect all fragments belonging to a given peak category to be identical in size. However, when the data are viewed in Genotyper, fragments from different samples belonging to a given peak category will almost always show variation in size. For example, one sample may have a fragment size of 76.4 bp, while a homologous fragment in another sample may be sized at 76.8 bp.

Labeling fragments. For a given fragment, homologous fragments across all samples may be identified if an appropriately sized "window" for that peak category is determined (Figure 2.1). It is possible to define such a window in Genotyper and automatically label fragments across all specimens that fall within the window with a peak category tag. However, this feature is often not useful for the following reasons:

First, variation in run conditions on the automated Genetic Analyzer may cause slight differences in fragment sizing. If the differences are large enough and fragment sizes fall outside the defined peak category window, fragments will be incorrectly labeled. These differences are most notable between batches of AFLP assays if the Genetic Analyzer capillary arrays are changed. This variation causes fragments to be consistently sized differently from previous batches, especially for fragments less than 100 bp in size.

Second, variation in size calling almost always results in some fragments being mislabeled as belonging to the next-higher or next-lower category, or not labeled because they fall outside the limits of the category window.

Third, users may remove an automatically generated label from a given fragment, but there is no facility in Genotyper to reassign the category that a fragment is automatically allocated to. Without this capability, users are left with the tedious and error prone task of noting down which fragments require reassignment, generating the absence-presence matrix, locating the particular '1' or '0' that corresponds to the fragment that requires reassignment, then changing the absence-presence matrix appropriately.

2.3.3 Challenges in Peak Height Analysis

In most cases, a given fragment for an individual is scored as either absent or present, depending on whether an AFLP fragment was observed at that locus. For a fragment that is present, Genotyper not only provides information about the size of that fragment, but also its "peak height". The peak height reflects the amount of PCR product of a

particular size. The amount of PCR product is likely to vary slightly across samples even under controlled laboratory conditions. This creates variation in peak heights. To ameliorate the effects of this variability, peak heights within each peak category are normalized prior to further analysis. The normalization is performed by using the "sum of signal" option in the "Calculate Scale Factors" function in Genotyper.

Delineating populations. After the peak heights are normalized, it is sometimes possible to assign individuals to a known population based on average peak heights (personal observation). For example, in a study involving two populations A and B with multiple individuals sampled from each population, a particular AFLP marker may exhibit a bi-modal distribution: individuals from population A may have lower average peak heights than individuals from population B. It is therefore possible, but unlikely, that N populations (sampled with multiple individuals per population) will exhibit a N-modal distribution. However, this has never been observed in practice: most peak categories exhibit randomly distributed peak heights, with only a few categories exhibiting bi-modal distributions. In the latter case, individuals from a given group of populations may be distinguished from individuals from the other populations from their average peak heights.

Identifying spurious fragments. AFLP PCR reactions may produce spurious fragments that appear as peaks with relatively low heights. For example, fragments with peak heights below 50 are almost always considered spurious fragments, and are scored 'absent'. In Genotyper, users are able to specify a threshold height below which such fragments are ignored. However, this is currently done in an *ad-hoc* manner. Moreover, Genotyper currently displays fragment peaks by their un-normalized heights, not the normalized heights, which is misleading if the user were trying to determine threshold values by eye-balling the AFLP profile. It would be desirable to have a capability to initially include all fragments and perform a *post-hoc* analysis of the distribution of peak heights to determine the threshold value for peaks to be considered present. Ideally, there should not only be threshold values to consider a fragment present or absent, there should

also be a threshold value to classify fragments as ambiguous. These ambiguous fragments would be scored with '?' instead of '1' or '0'. The '?' convention for missing data is currently adopted by PAUP (Swofford 1998), Arlequin (Schneider et al. 2000) and other phylogenetic analysis programs.

There is currently no such capability to analyze the distribution of peak heights, either for the purposes of identifying peak categories which are associated with certain populations, or for identifying spurious AFLP fragments.

2.3.4 Solution – AFLPal and AFLPexel

These challenges are addressed with a custom-written program for the PC, called AFLPal. AFLPal is particularly useful for situations where issues related to peak height analysis highlighted in the previous section are frequent.

Unfortunately, even though AFLPal reduces the amount of subjectivity in determining these threshold values, it does not completely eliminate all the subjectivity. Part of the reason is the lack of a theoretical framework linking AFLP fragment copy numbers to population genetic processes. The lack of such a framework precludes an *a-priori* determination of threshold values which can be consistently applied across each AFLP loci. Furthermore, as noted above, peak heights are likely to be affected by PCR conditions in a manner which may not be predictable.

Personal experience has also shown that average peak heights are affected by the type of pre-selective primer used for the AFLP PCR. For example, with *Euphydryas editha*, AFLP fragments generated using the small genome pre-selective primer exhibited lower average peak heights than if the regular genome pre-selective primer were used. In the former case, fragments with peak heights below 50 were disregarded because they likely represented spurious fragments. Peak heights of 500 were considered exceptional, while the majority of fragments exhibited peak heights less than 200. On the other hand, with

the regular genome pre-selective primer, fragments with peak heights below 100 were typically discarded, and peak heights between 500 and 2500 were common (see Figure 2.3 for examples). This argues against using a simple rule of thumb for setting threshold values, such as "discard any fragment with peak height less than 10% of the tallest peak for that locus". The use of such a heuristic would eliminate viable fragments from loci with a relatively large range of peak heights.

If the user determines that the AFLP data can be scored directly in Genotyper without analyzing the distribution of peak heights, AFLPexel should be used instead. This is described briefly in Section 2.5 "AFLPexel".

2.4 AFLPal

2.4.1 Overview

To obtain an absence-presence matrix from the AFLP data produced by the Genetic Analyzer, the data is first analyzed in Genotyper, then in AFLPal. This process is briefly outlined below.

Analysis in Genotyper. The Genetic Analyzer saves raw fragment data into files ending with the .FSA extension. These files are imported into Genotyper with a special Genotyper template file created for use in conjunction with AFLPal and AFLPexel. This template file contains a number of macros and specially defined peak categories. A macro is a series of pre-defined procedures within Genotyper which are sequentially executed once the macro is invoked. A specific template file exists for each color dye: for example, specimens assayed with the FAM-dye labeled primers should be analyzed using the template file specifically designed for the FAM-dye. After the data have been imported, the peak heights are normalized and the user proceeds to define peak categories. Recall that peak categories are synonymous with AFLP loci. Although monomorphic AFLP loci are not informative in character-based parsimony analysis, they are

nevertheless informative in distance-based analyses. Once the peak categories have been defined, this is followed by automated labeling of the peaks. The user next edits the automatically-labeled peaks to remove or reassign peaks labels which have been incorrectly assigned. Once the peak labels have been edited, the user invokes the macro contained in the template to export the labeled peaks to a tab delimited text file that can be read by AFLPal.

Analysis in AFLPal. AFLPal converts the text export file into a Microsoft Access database file and displays the frequency distribution for each peak category. The user is able to specify, for each AFLP marker, threshold values for classification of peaks into '1', '0', and '?' based on the peak height. From this information, AFLPal generates a tab-delimited text file which may be read by phylogenetic programs.

The following two sections briefly explain how certain features in Genotyper may be used to efficiently prepare data that can be analyzed with AFLPal.

2.4.2 Fragment Categorization using Genotyper

One way to create peak categories in Genotyper is to first perform an automatic labeling of all peaks, followed by invoking the function "*Make from labels*". This is not recommended. Instead, the peak categories should be created manually. This is easily achieved by setting the plot options for the upper pane of the plot window to display "overlapped dye/lanes" instead of "separate dye/lanes". This results in sample profiles being plotted on top of each other (Figure 2.1). This allows the user to gauge the suitability of fragments as diagnostic loci. For example, a peak category with a moderately high frequency of overlapping peaks (for example, P124 in Figure 2.1) may be a more reliable locus than one with an extremely low frequency of overlapping peaks (for example, P127 in Figure 2.1). The latter may be rare alleles – which are useful for certain analyses – or they may be PCR artifacts. Due to the anonymous nature of AFLP markers, it is recommended that these peaks be treated conservatively as PCR artifacts.

Moreover, in a typical AFLP study, hundreds of markers are employed. It is more conservative to eliminate from analysis potentially suspicious markers. This allows subsequent analyses to be based only on markers that can be scored with a high degree of certainty. Slightly suspicious markers may however be useful in cases if they are found to be fixed for a certain population with a limited number of sampled specimens.

Experience also suggests that a category should be defined with a window not exceeding ± 0.4 bp of the mean fragment size. Once all categories have been defined, the Genotyper "*Label Peaks*" function should be invoked to automatically label the peaks in each sample. AFLPal requires that each peak be labeled with its size, its height divided by the normalization scale factor, the category's name, the label / peak source, and the peak modulation score.

Once the peaks are automatically labeled by Genotyper, the user must verify the labeling of each peak to correct for mislabeling due to variations in PCR and run conditions between samples. Prior to editing, the "*Set Click Options*" function should be invoked to label each peak with its size, its height divided by the normalization scale factor, the requested text, the label / peak source, and the peak modulation score. Clicking on an existing peak label erases that label. Clicking on an unlabeled peak allows the user to name the peak label by typing in the desired category name.

At this stage of the analysis process, users have been known to try to identify spurious peaks based on the relative height of each peak within a given peak category (see Section 2.3.3 "Challenges in peak height analysis"). That is, users may be tempted to eliminate from further consideration peaks with relatively low peak heights which they suspect to be spurious PCR artifacts. This is not recommended. All peaks should be labeled regardless of their peak heights. The decision to classify peaks with low heights as absent should be delayed until the distribution of peak heights for a given peak category can be assessed in AFLPal.

2.4.3 Peak Height Analysis using AFLPal

After the data are edited in Genotyper, the macro included with the AFLPal Genotyper template file named "AUT: Make Excel Data" should be executed. This macro transfers the information from the labeled peaks into a table which can then be exported into a tab-delimited text file. This file is used by AFLPal and AFLPexel for further analysis.

AFLPal allows the user to browse through the histogram plot for each peak category (Figure 2.2). Slider bars allow the user to position an indicator along the length of the histogram plot to indicate the threshold peak heights for labeling peaks as '0', '1', or '?'. On the other hand, users may choose not to score fragments with '?', electing instead to score fragments as strictly '0' (absent) or '1' (present). The challenges in creating an objective standard for peak height threshold values are outlined in section 2.3.4. Instead of attempting to create such a standard, examples of how threshold values are determined and the rationale behind the process are described below.

Figure 2.3 shows four AFLPal histograms, each corresponding to a peak category. The histograms are calculated using 501 samples. Each histogram shows the frequency distribution of samples which have labeled peaks (see prior section). Each labeled peak represents a putative fragment which is to be scored '1'. Samples without labeled peaks, which are not shown on the histogram, are automatically scored '0'. Prior experience with this data set suggests that the peak height threshold for 'present' should be set at 100.

Figure 2.3A shows a locus with 357 labeled peaks. To determine if all the labeled peaks should be scored '1', we examine the second frequency interval. It shows that there are seven fragments with heights between 56 and 112. The heights of these seven fragments are: 70, 75, 78, 79, 80, 104, and 111. If we apply the pre-established threshold value of 100, the first five of the seven fragments would be scored '0', and the others would be scored '1'. However, the histogram suggests that peak heights are distributed normally

with no discernable bi-modal distribution. The threshold height of 100 was not applied in this case, and all the samples with labeled peaks were scored '1'.

Figure 2.3B shows an AFLP locus with 10 labeled peaks, all taller than the threshold height of 100. Due to the small number of labeled peaks, it is difficult to discern whether the data fits any distribution. In this case, there is no reason to score any of the labeled peaks '0', so all 10 labeled peaks were scored '1'.

In Figure 2.3C, there is one labeled peak with a height less than the threshold value of 100. This sample corresponding to this labeled peak was scored '0'. The other seven samples with labeled peaks were scored '1'.

Figure 2.3D shows a locus with the majority of labeled peaks exhibiting heights of more than 823. These peaks were scored '1'. There are two labeled peaks in the height range 449 – 524 found between two large gaps in the distribution. These peaks were scored '?', since they were too big to be considered spurious AFLP fragments, but yet seem isolated from the main group of fragments that are higher than 823 in height. The rest of the labeled peaks were scored '0'.

The final step involves invoking the function to export the AFLPal data to a tab delimited text file for further analysis in other programs.

2.5 AFLPexel

AFLPexel is a Visual Basic for Applications (VBA) macro written in Excel. It converts the Genotyper export file into an Excel spreadsheet that can then be exported for use in phylogenetic analysis programs. Its main difference from AFLPal is that AFLPexel does not offer any capability to perform peak height analysis. It is therefore used in situations where the issues highlighted in Section 2.3.3 ("Challenges in peak height analysis") are not apparent in the data.

AFLPexel processes the Genotyper data and lists, for each sample, the number and names of all the peaks associated with that sample. For each AFLP locus, AFLPexel reports the mean peak height as well as the smallest and largest peak heights and the sample IDs associated with those peaks. It also reports the percentage of fragments scored 'present' for each locus. An absence-presence matrix that represents the absence or presence of a given fragment for each individual sample ID is also created as a separate worksheet.

To use AFLPexel, the raw AFLP data from the Genetic Analyzer must first be analyzed as described in Sections 2.4.1 and 2.4.2, above.

Figure 2.1: Partial screen capture of Genotyper v3.6.

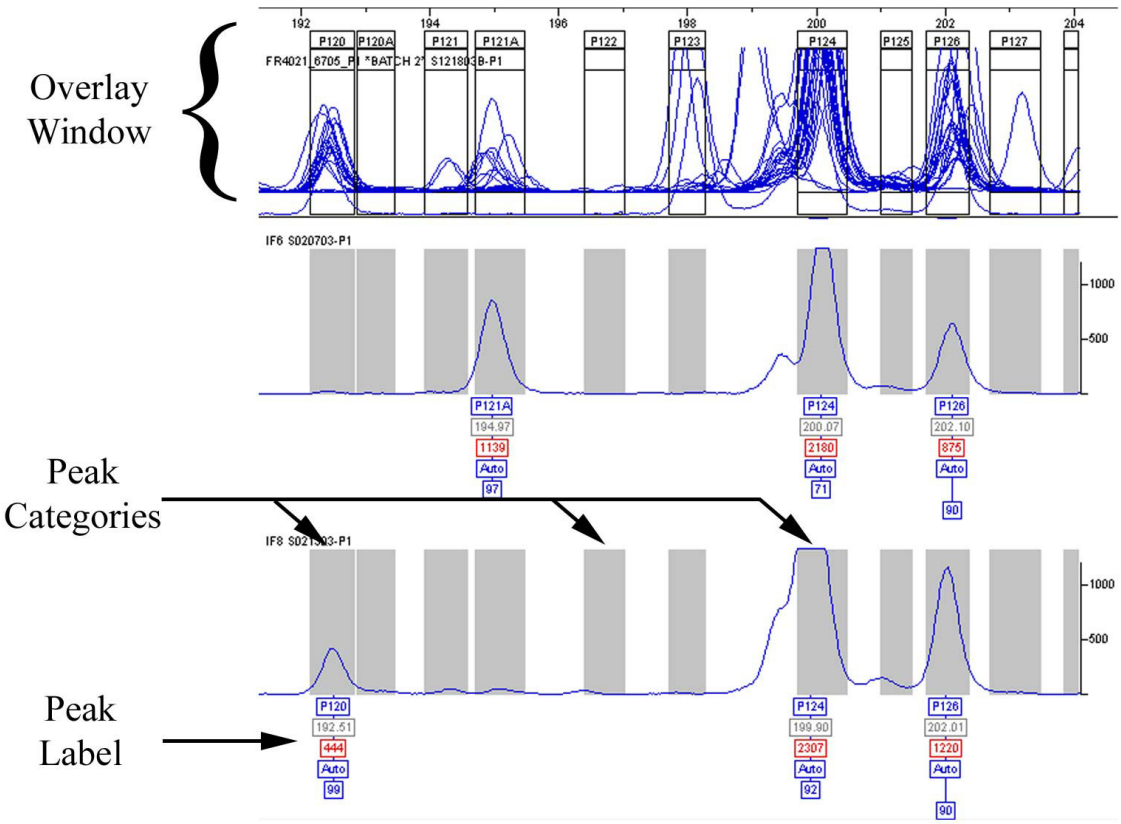


Figure 2.2: Partial screen capture of AFLPal.

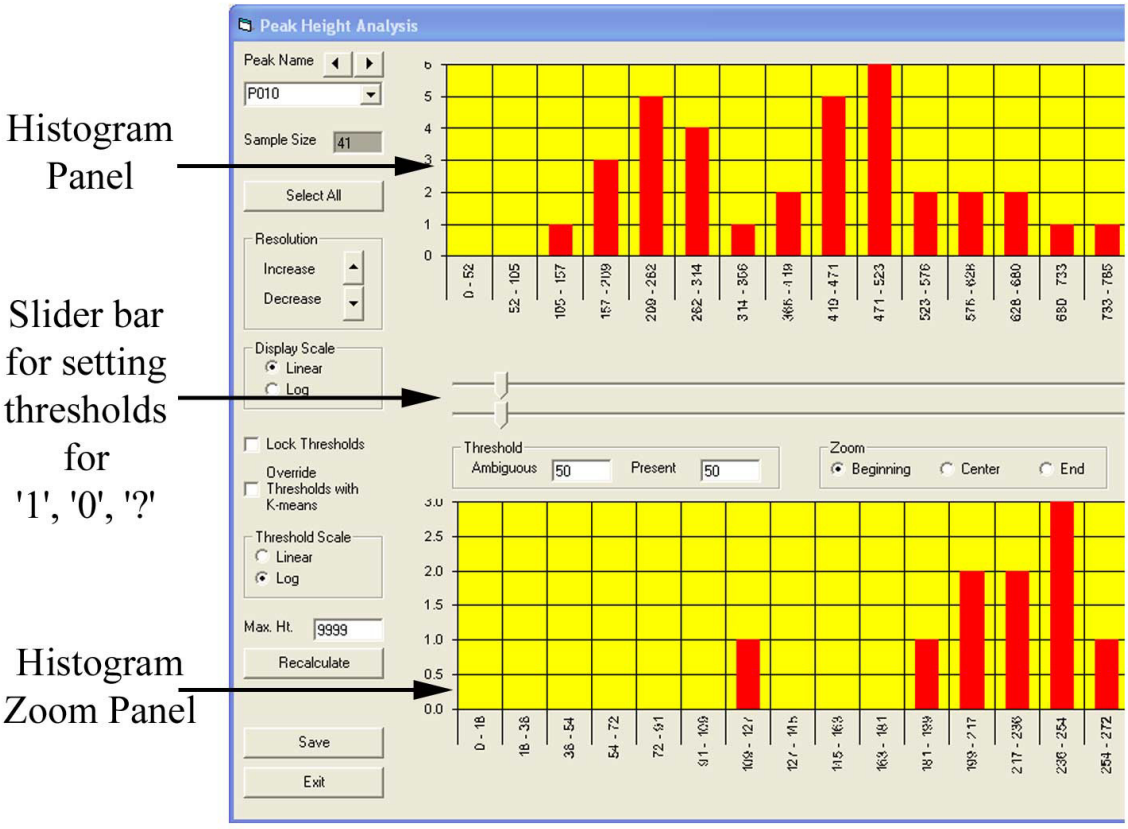
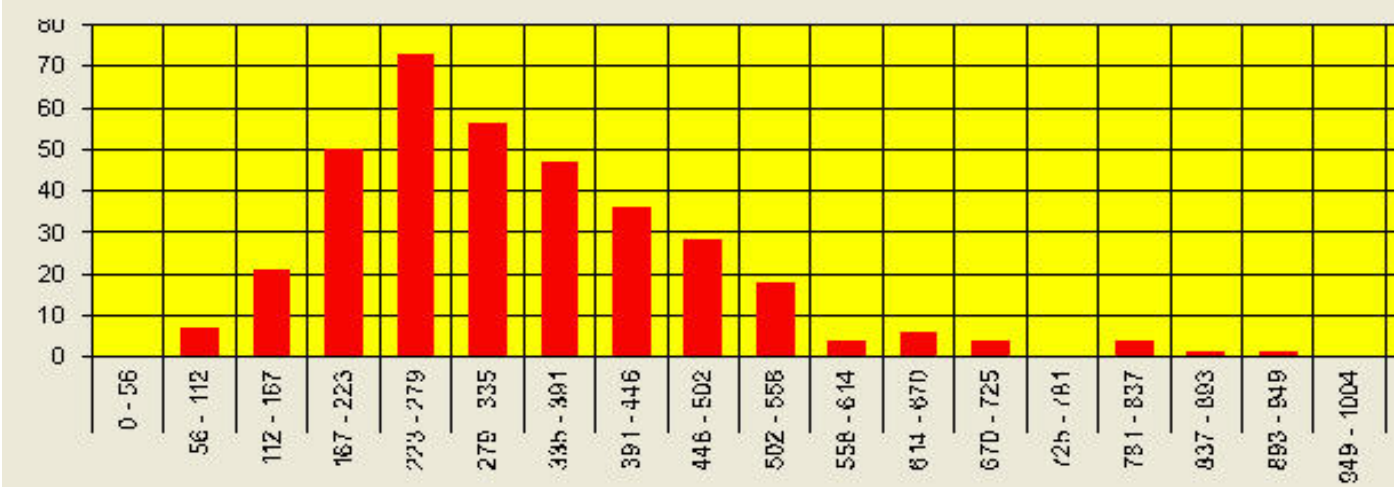


Figure 2.3: **AFLPal histograms demonstrating how fragments are scored. *PH* denotes fragment peak height. Total sample size is 501. (A) *PH* ≥ 56 scored as '1', '0' otherwise (N = 357). (B) *PH* ≥ 325 scored as '1', '0' otherwise (N = 10).**

(A)



(B)

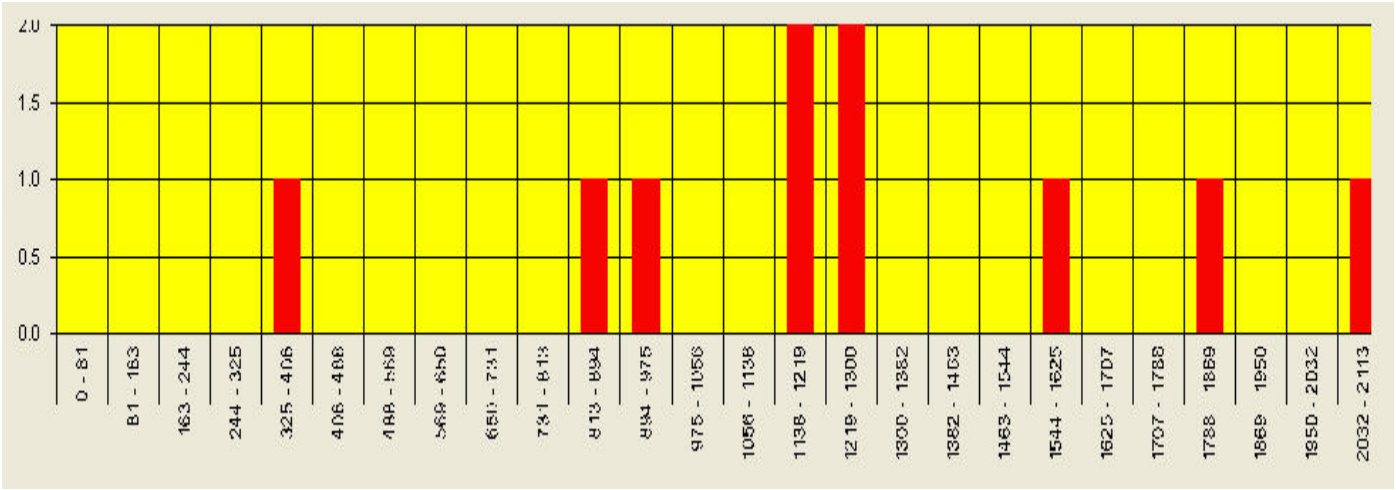
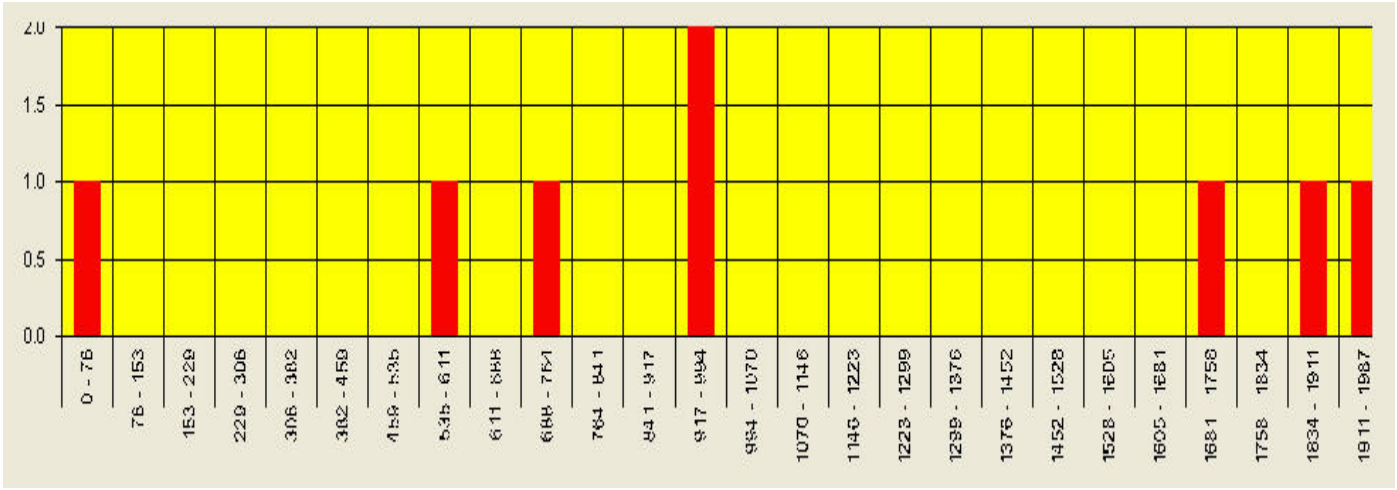
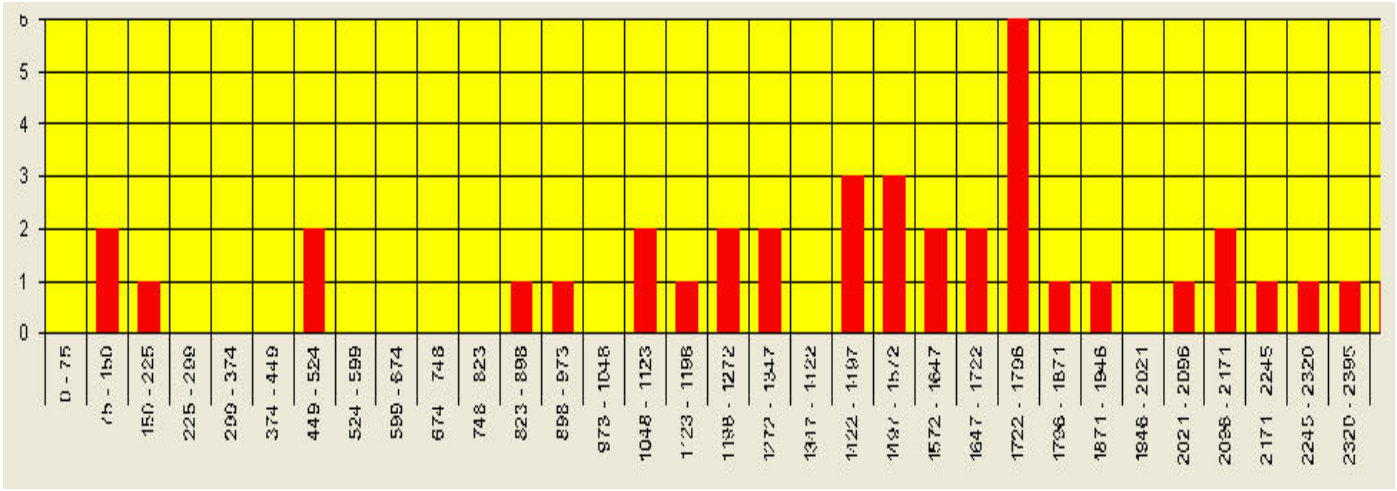


Figure 2.3: **(C) $PH \leq 78$ scored as '0', $PH \geq 535$ scored as '1' (N = 8). (D) $PH \leq 299$ scored as '0', $300 \leq PH \leq 674$ scored as '?', $PH \geq 675$ scored as '1', '0' otherwise (N = 39).**

(C)



(D)



References

Jones, C. J., K. J. Edwards, et al. (1997). "Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories." Molecular Breeding **3**(5): 381-390.

Mueller, U. G. and L. Wolfenbarger (1999). "AFLP genotyping and fingerprinting." Trends in Ecology and Evolution **14**: 389-393.

Schneider, S., D. Roessli, et al. (2000). Arlequin: A software for population genetics data analysis., Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts., Sinauer Associates.

Chapter 3: Host-Associated Genetic Differentiation in Allopatric Populations of the Marsh Fritillary *Euphydryas aurinia*

Abstract. I report a case where the degree of genetic differentiation between conspecific herbivorous butterflies on different host species depends on whether the hosts are sympatric or allopatric. Host-associated genetic differentiation exists among allopatric populations of the insect but not within a single population that uses two hosts. The most relevant context is the current resurgence of interest in host-associated speciation in herbivorous insects. The most cited examples include *Rhagoletis pomonella* on apple and hawthorn, *Acyrthosiphon pisum* on alfalfa and clover and *Eurosta solidaginis* on *Solidago altissima* and *S. gigantea*. These studies have all found host-associated genetic differentiation between sympatric "host races." They have also recorded host-associated selection that promotes such differentiation and various classes of pre-mating and post-mating isolation that help maintain it. Because these studies have deliberately concentrated on species where host-associated isolating mechanisms are evident, such as mating on hosts, it is now pertinent to ask whether and under what conditions host-associated genetic differentiation occurs in species where such mechanisms are weak or absent. Here, I use AFLP markers to study such an insect, the butterfly *Euphydryas aurinia.* Four of the populations studied were monophagous on *Succisa*, six were monophagous on *Lonicera*, and one population utilized both hosts. There was substantial overall genetic differentiation between populations ($\Phi_{ST} = 0.18$, $P < 0.001$). Geographical isolation was partially responsible for genetic differentiation ($r = 0.34$, $P = 0.03$). Strong effects of host association on genetic differentiation were found between monophagous populations on *Succisa* and *Lonicera* ($F_{CT} = 0.12$, $P < 0.001$). This finding was reinforced by the significant correlation between population pairwise genetic differentiation and two measures of host utilization (Jaccard coefficient: $r = 0.58$, $P = 0.001$ and Sorensen coefficient: $r = 0.60$, $P < 0.001$) after controlling for effects of geography. At the dual host-use population, there was no significant genetic

differentiation between larvae found on *Succisa* and those found on *Lonicera* (F_{ST} = 0.04, P = 0.11). In this case where there is no trend towards host-associated mating, the effects of host-use on gene flow appear insufficient to allow sympatric host-associated genetic differentiation but sufficient to allow allopatric differentiation. In herbivorous insects as in other systems, allopatry assists differentiation.

3.1 INTRODUCTION

3.1.1 Background

Local adaptation may disrupt the expected association between gene flow and distance by acting as a barrier to gene flow through behavioral traits (Kawecki 1998; Berlocher 1999) or physiological traits. Such effects are expected to be most pronounced in organisms showing strong local adaptation, such as herbivorous insects, which often show a high degree of local host specialization (Mopper and Strauss 1998). It follows that the use of different diets by allopatric populations of herbivorous insects may contribute significantly to their genetic isolation. Here, I test this hypothesis of host-associated genetic differentiation and ask whether and how it interacts with effects of geographic isolation.

Host-associated genetic differentiation of herbivorous insects has been attributed to host fidelity, defined by Feder (1998) as the tendency of insects to mate and oviposit on the same host plant species on which they themselves developed (Feder et al. 1993; Feder et al. 1994; Via 1999; Abrahamson et al. 2001). In addition to host fidelity, Funk et al. (2002) stress the role of two other premating barriers in ecologically driven speciation: temporal isolation associated with differences in timing of host plant development (Wood and Keese 1990; Feder et al. 1993; Feder and Filchak 1999; Thomas et al. 2003), and behavioral isolation associated with courtship signals or mating preferences (Abrahamson et al. 2001; Thomas et al. 2003).

In addition to premating isolation mechanisms, Dres and Mallet (2001) included in their definition of "host race" the requirement that hybrids of the parental types demonstrate lower fitness than their parents on their respective hosts. Negative performance correlations are often the final ingredients in most sympatric speciation models since they penalize failure to maintain host fidelity (Berlocher and Feder 2002). Such effects may be intrinsic (genetic incompatibilities) or extrinsic (ecologically based) (Via et al. 2000; Berlocher and Feder 2002). Examples of hybrid inviability include the gall fly *Eurosta solidaginis*, where hybrids of the host races on *Solidago altissima* and *S. gigantea* perform more poorly than parental types on all host plants (Craig et al. 1997; Itami et al. 1997; Craig et al. 2000; Craig et al. 2001). Other examples of negative performance correlations across hosts include hybrids of *Acyrthosiphon pisum* (Via et al. 2000) and *Rhagoletis pomonella* (Filchak et al. 2000).

The speciation studies detailed above have all found host-associated genetic differentiation between host races. Each of these study systems exemplifies more than one of the categories of pre-mating and post-mating isolation that are thought to promote or maintain such differentiation in sympatry. For example, *Rhagoletis* flies show host fidelity in choice of mating site, mating on the host species on which they developed. In addition, there is phenological isolation between flies on different hosts driven by differences in the time of host fruit ripening. Here, I study a set of allopatric populations that broadly lack the phenomena (see below) that are generally thought to be responsible for host-associated genetic differentiation in the systems listed above.

3.1.2 Study System

E. aurinia is a univoltine species found throughout Europe from Fennoscandia to south Greece (Joyce and Pullin 2001). Populations are largely monophagous: throughout its range. Primary host plants are *Succisa*, *Knautia*, *Scabiosa*, *Cephalaria* (Dipsacaceae), *Lonicera* (Caprifoliaceae) and *Gentiana* (Gentianaceae) (Singer et al. 2002). It is monophagous on *Succisa pratensis* in northern Europe (Wahlberg et al. 2002).

Population structure is variable: in some regions *E. aurinia* exists as discrete, well-defined populations, while in others it forms large metapopulations that have been subjects of studies in metapopulation dynamics (Wahlberg et al. 2002; Wahlberg et al. 2002; Anthes et al. 2003).

3.2 MATERIALS AND METHODS

3.2.1 Sampling

Larvae were collected from 11 locations, three in southern France, and eight in north-eastern Spain (Figure 3.1). Four of these populations were monophagous on *Succisa*, six were monophagous on *Lonicera*, and one population utilized both hosts. No *Succisa* was present at the *Lonicera*-feeding populations, which occurred in dry Mediterranean scrub habitats or in open woodland. The *Succisa*-feeding populations were in humid meadows where *Lonicera* did not grow, but at two sites (COU and CJ; see Figure 3.1) *Lonicera* grew in surrounding woodland where it was not used by the butterflies. The site where both hosts were used, LA, was a single meadow about 40m by 90m with larvae on *Succisa*, surrounded by dry scrub with larvae on *Lonicera*. Although there was no *Lonicera* in the meadow, larvae on *Lonicera* were found within 5m of larvae on *Succisa*. No larva was gathered on either host more than 50m from the nearest larvae on the other host. At the time of sampling, which was conducted in the course of a single day, larvae were abundant on both hosts. They were all at a sufficiently early stage that they would still have been on the host chosen by their mother for oviposition, and would not have been able to move to the alternate host plant species.

The populations examined in this study and the respective sample sizes are listed in Table 3.1, and their geographical relationship is depicted in Figure 3.1. The latitudinal band between FORN and LA represents a region where both *Succisa* and *Lonicera* are utilized: there is no known utilization of *Lonicera* for oviposition north of FORN, and no known utilization of *Succisa* south of LA.

Both adult insects and larvae were used for the molecular assay. Larvae were collected from the field between 2001 and 2003. At LA where both *Lonicera* and *Succisa* were used, an equal number of larval specimens were collect from each host. These specimens were labeled with the host plants brought back live to the laboratory, where two from each family were frozen at -80 °C. Freshly dead or dying adults were similarly preserved.

3.2.2 AFLP Assay

Sample preparation. DNA was isolated from specimens either frozen at -80 °C or preserved in 100% EtOH. For adult specimens, a single leg was removed for DNA isolation, and the rest of the body stored for future use. In order to minimize possible contamination from parasitoids in the larval body, DNA was extracted from the heads of larvae.

DNA extraction. Qiagen DNeasy Tissue kits (Catalog No. 69506) were used for DNA extraction. Instead of using Qiagen's tissue lysis buffer, a CTAB extraction buffer (10% 1 M Tris-HCL pH 8.0, 28% 5 M NaCl, 4% 0.5 M EDTA pH 8.0, 2% CTAB) was used instead. Prior to extraction, 2 µl of beta-mercaptoethanol per 1ml of extraction buffer was added to the buffer. Samples were crushed in a 1.5ml Eppendorf tube in liquid nitrogen, and then homogenized in 200ul of the extraction buffer. The Qiagen protocol was followed for the rest of the steps. The DNA was eluted in two 100 µl washes of buffer AE for a final DNA volume of 200µl.

AFLP PCR reaction. The Applied Biosystems (ABI) AFLP protocol for small plant genomes (www.appliedbiosystems.com; protocol 4303146) was used with the following six modifications: (A) 30 ng of DNA was used for restriction-ligation and incubated using a thermocycler with a heated lid for 3 hours. (B) 120 µl of $TE_{0.1}$ was added to the restriction-ligation product instead of the recommended 189 µl. (C) Pre-selective amplification with the regular genome preselective primer mix was next performed in 10 µl reactions. (D) Instead of the recommended 20 cycles of denature – anneal – extension,

25 cycles were used instead. (E) The pre-selective product was not diluted for the selective amplification. (F) For the selective PCR, the 25 cycles of denature – anneal – extension (20 s at 94 °C, 30 s at 56 °C, 2 min at 72 °C) were used instead of the recommended 20.

The four primer pairs from ABI used for this study are: (1) EcoRI –ACA / MseI – CTG, (2) EcoRI – ACA / MseI – CAG, (3) EcoRI – AAG / MseI – CTG, and (4) EcoRI – AAG Joe / MseI – CAG. 3 µl of the selective product was mixed with 0.4 µl of GeneScan 500 Rox and 6.6 µl of Hi-Di Formamide. Samples were processed on an automated ABI 3100 Genetic Analyzer.

3.2.3 Analysis

Scoring of AFLP data. Raw fragment data from the 3100 Genetic Analyzer were scored in Genotyper v3.6 (PE-Biosystems). A custom written Genotyper macro was used to export the scored data to a custom written PC program, AFLPal (see Chapter 2 "An efficient AFLP protocol for butterflies: from laboratory to analysis"). In most cases, a given locus for an individual is scored as either absent or present, depending on whether an AFLP fragment was generated for that locus. However, certain populations were characterized not only by the absence or presence of a fragment, but also by the number of copies of that fragment. This was reflected by the fragment peak height in the Genotyper software. Peak heights with bimodal frequency distribution reflect individuals from two different populations and should be scored as such (personal observation). AFLPal was used to analyze the frequency distribution of the number of fragments associated with any one locus. Based on that analysis, the program allows the user to control how the absence-presence matrix is to be generated. In addition, AFLPal allows the user to specify parameters to denote a fragment that is present (scored as '1'), absent (scored as '0'), or ambiguous (scored as '?'). The latter state can be used by many programs to denote missing data.

Genetic analysis. Arlequin (Schneider et al. 2000) was used to analyze the absence – presence matrix generated by AFLPal. The AMOVA (Excoffier et al. 1992) module in Arlequin was used to determine Φ statistics, equivalent to the Weir and Cockham's (1984) θ statistics (Excoffier 2001). All Φ statistics were computed for statistical significance with 5040 permutations. AMOVA was also used to quantify the amount of variation between the two groups of populations: populations on *Lonicera* and populations on *Succisa*. LA, the dual host-use population, was excluded from the AMOVA analysis. The population pairwise Φ_{ST} matrix calculated by Arlequin was used in full and partial Mantel correspondence analyses to calculate r, the Pearson correlation coefficient, using the program ZT (Bonnet and Van de Peer 2002). All significance tests with ZT were performed with 10000 permutations. To test for effects of IBD, the correspondence test was done between the Φ_{ST} matrix and a pairwise geographical distance matrix between populations. The program Range (Luetgert, USGS) was used to calculate the geographical separation between populations measured as straight line distances between two points.

Host utilization matrix. In order to perform correlates of genetic distances with differences in host-use, a host utilization matrix was created for use in Mantel correspondence tests. This matrix summarizes the differences in host utilization between populations. To create this matrix, an Excel spreadsheet with the populations listed as rows and host plants listed as columns was created (Table 3.1). For a given population, '1's were scored in the relevant cells to represent utilization of a particular host for oviposition. '0's were scored when that particular host was not utilized. PopTools (Hood 2003), an Excel plug-in utility for the analysis of matrix population models, was then used to generate two host-utilization distance matrices: one using the Jaccard coefficient, and the other the Sorenson coefficient. Each cell in these distance matrices represents the degree of difference in host utilization for oviposition between populations.

3.3 RESULTS

A total of 89 specimens were used for the AFLP assay: 38 adults and 51 larvae. The primer pairs EcoRI –ACA / MseI – CTG, EcoRI – ACA / MseI – CAG, EcoRI – AAG / MseI – CTG, and EcoRI – AAG Joe / MseI – CAG generated 64, 74, 106, and 74 markers respectively. A total of 318 AFLP loci were identified, all polymorphic. Table 3.2 shows the pairwise Φ_{ST} matrix generated by Arlequin for the populations at the regional scale.

Figure 3.2 shows the results of the NMDS analysis performed on the population pairwise distance matrix. Table 3.4A shows the first ten eigenvalues and cumulative percentage explained by each ordination axis. Axis #1 accounts for 49% of the total variation, while axis #2 accounts for a further 28%, for a total of 77% of the total genetic variation captured in the two-dimensional NMDS map.

Host-associated structure. Overall Φ_{ST} was estimated at 0.18 ($P < 0.001$), with 82% of the total genetic variation found within populations. Populations which utilized *Succisa*, namely – RED, CJ, COU, and LNB – were found on the positive side of Axis #1, while populations which utilized *Lonicera* – TOR, ELG, COL, FORN, DAR, and MC – were found on the negative side of the axis. The population LA, which uses both *Lonicera* and *Succisa*, was found near the origin of Axis #1 (NMDS coordinates: Axis #1 = -0.03, Axis #2 = 0.13).

An AMOVA analysis found a significant degree of genetic differentiation between the groups of populations on *Lonicera* and on *Succisa* ($F_{CT} = 0.12$, $P < 0.001$) (Table 3.5). Mantel correspondence tests of the population pairwise Φ_{ST} matrix against the host utilization matrix support the finding of host association for both the Jaccard coefficient matrix ($r = 0.55$, $P < 0.001$) and the Sorensen coefficient matrix ($r = 0.58$, $P < 0.001$). A partial Mantel test controlling for the effects of distance increases the correlations slightly

($r = 0.58$, $P = 0.001$ and $r = 0.60$, $P < 0.001$ for the Jaccard and Sorensen matrices respectively).

Geography associated structure. The position of populations on the NMDS map (Figure 3.2), in addition to being associated with their host utilization, also bears a strong geographical signal. The *Lonicera* populations FORN, DAR, and MC form an arc on the population map (Figure 3.1) stretching from north to south. These three populations are clustered on the lower left quadrant of the NMDS map. The other three *Lonicera* populations, ELG, TOR, and COL are found on the southern end of the range of this study, and they are clustered in the upper left quadrant of the NMDS map. The *Succisa* populations COU, CJ, and LNB are found in the center of the study range, and cluster together on the lower right quadrant of the NMDS map.

Isolation by distance. The minimum, maximum, and mean distances between populations, measured as the crow flies, are 14.1km, 145.5km, and 68.6km respectively. Figure 3.4 shows a graph of pairwise geographical distances plotted against pairwise Φ_{ST} values. A moderate slope is apparent, but there appeared to be a lot of scatter around the regression line ($R^2 = 0.08$). The Mantel correspondence test did not suggest genetic differentiation mediated by geographical distance ($r = 0.29$, $P = 0.06$). However, when the effects of host association were held constant, there was a significant effect of distance on genetic differentiation ($r = 0.34$, $P = 0.03$).

Dual host population. The NMDS solution for individuals from the dual host-use population LA is shown in Figure 3.3, and Table 3.4B shows the first ten eigenvalues and cumulative percentage explained by each ordination axis. There was no significant genetic differentiation between larvae found on *Succisa* and those found on *Lonicera* ($F_{ST} = 0.04$, $P = 0.11$, Table 3.5B).

3.4 DISCUSSION

3.4.1 Genetic Structure

Allopatric *E. aurinia* populations in this study separated by an average of 68.6km and a study area covering 10, 000km^2 demonstrated high differentiation with a fixation index of 0.18. Wright (1978) suggested that F_{ST} values of 0.15 to 0.25 indicate high genetic differentiation. This amount of genetic differentiation is similar to that of 0.19 found for *Euphydryas editha* in the Sierra Nevada on the western coast of the United States (Chapter 5). Effects of IBD were found after controlling for the effects of host association. This contrasts to an allozyme study of *E. aurinia* populations in the United Kingdom by Joyce and Pullin (2001). *E. aurinia* in Ireland and the United Kingdom utilize mainly *Succisa pratensis* (Feehan and O'Donovan 1996; Asher et al. 2001). In that study, they found no evidence for IBD using allozyme data from 11 populations collected over an approximate area of 180,000km^2.

The detection of IBD in this study but not in the Joyce and Pullin (2001) study may be due to region-wide adaptations which are different in the United Kingdom and the current study area. It could also be due in part to the difference in genetic resolution afforded by allozymes and AFLP markers. Allozymes, although offering the advantages of co-dominance, are unlikely to show high levels of variation, as shown for other butterflies (Saccheri et al. 1999). This may lead to an underestimate of the level of differentiation compared to a higher resolution marker when analyzed over a similar geographical scale. Allozyme markers have been also shown under certain cases to be under the influence of natural selection (Anderson and Oakeshott 1984; Begun and Aquadro 1994; Hedrick 1999) and thus not suitable as neutral molecular markers. AFLP studies typically employ a large number of markers. These markers are randomly scattered throughout the genome, which helps to decrease the chances that a high proportion of markers are under selection.

3.4.2 Host Association

There has been a recent resurgence of interest in sympatric speciation in herbivorous insects. The most cited example is that of the host races of *Rhagoletis pomonella* on apple and hawthorn (Bush 1969; Bush and Smith 1998; Feder 1998). In that system, it is clear that members of each host race tend to return to their natal host plant for mating and oviposition. Strong host fidelity is also exhibited in the host races of the pea aphid *Acyrthosiphon pisum* which utilize either alfalfa or clover (Via 1991; Via 1991; Via 1999). Upon landing on a potential host plant, adults will first sample the plant tissues before accepting or rejecting a plant (Caillaud and Via 2000). If the adult discovers that it is a non-host, it will refuse to feed and starve to death if prevented from leaving (Del Campo et al. 2003).

The level of host-associated genetic differentiation between groups of populations on either hosts found in this study, $F_{CT} = 0.12$, is substantial, comparable in magnitude to the overall fixation index of $\Phi_{ST} = 0.18$. Mantel tests between the population pairwise Φ_{ST} matrix and two similar measures of host association suggests a significant relationship between genetic differentiation and host utilization (Jaccard coefficient matrix: $r = 0.58$, $P = 0.001$ and Sorensen coefficient matrix: $r = 0.60$, $P < 0.001$). This suggests that host association may constitute a potential barrier to gene flow between allopatric *E. aurinia* populations. In contrast, no genetic differentiation could be found between larvae gathered on *Lonicera* and *Succisa* at the biphagous site LA.

In the introduction I hypothesized that *E. aurinia*, in contrast to other herbivorous insects that have so far been studied in the speciation context, exhibits little evidence of host-associated isolation. Strong host-associated differentiation among allopatric populations has nonetheless emerged in *E. aurinia*. What is the nature of the evidence that host-associated isolation is weak?

Post-alighting oviposition preference and larval performance. Mazel (1986) exposed adult female *E. aurinia* to different hosts and found that *Succisa* was generally preferred, regardless of whether it was a locally available host. Singer et al. (2002) tested the preference of insects from two of the *Lonicera*-feeding populations used in the present study, MC and ELG. They confirmed Mazel's result that females tended to prefer *Succisa* over *Lonicera* for oviposition, but the most-preferred individual *Lonicera* ranked approximately equal with *Succisa*. In the light of this information, it is not surprising that insects from the *Succisa*-feeding populations RED and COU also preferred *Succisa* over *Lonicera*. However, almost all such insects accepted *Lonicera* for oviposition if they did not encounter *Succisa*. Larvae from *Lonicera*-feeding populations MC, ELG and FORN and from the *Succisa*-feeding population COU all grew well, survived well, and produced fertile adults on both hosts.

Absence of host-associated mate search. Habitat patches were surveyed at DAR. In these patches, the vegetation height and spatial arrangement of shrubs and trees is similar to those favored by males when choosing territories to defend. No patch contained two males for long, so some patches had a single territorial male and some had none. The presence/absence of the host, *Lonicera*, was also noted in each patch. The presence/absence of *Lonicera* was not significantly associated with the presence/absence of a territorial male (Table 3.6) ($P = 0.84$, Fisher's exact test) (M.C. Singer, unpublished).

Ease of mating and fertility of crosses between insects from different hosts. No obstacles to mating were observed in captive insects derived from populations on different hosts. Eggs laid as a result of crosses between insects associated with different hosts were completely fertile, and the larvae were able to develop on either host (M.C. Singer & B. Wee, unpublished data). The hybrid adult insects appeared healthy. Experiments to generate backcrosses and F2 lines are under way.

Phenological isolation. The only known cause of isolation in *E. aurinia* is a moderate trend for phenological separation: the time period in which insects are available for

mating is not identical in populations on the two hosts, though it is overlapping. A field comparison between two nearby populations at similar elevation on different hosts showed a mean difference in developmental stage of about ten days' growth (M.C. Singer, unpublished data). In many insects, including both *Rhagoletis pomonella* (Feder 1998) and *Euphydryas editha*, a close relative of *E. aurinia* (Boughton 1999), such a difference would be crucial to survival. However, in contrast to those of *R. pomonella* and *E. editha*, host plants of *E. aurinia* are apparently available for extended time windows, much longer than necessary for insect development. Feeding of wild-gathered hosts to captive larvae indicated that this apparent availability is real: both *Succisa* and *Lonicera* remained nutritionally suitable for young larvae after the local larvae had stopped feeding and entered diapause. Therefore, the difference in peak flight season between insects on the two hosts does not seem to render hosts phenologically unavailable to migrants between populations on different hosts.

3.4.3 A Topic in Need of Research: Pre-Alighting Oviposition Preference

At present there is no information on host searching behavior by female *E. aurinia* in the set of populations studied here, nor on the potential for local adaptation in this component of behavior to comprise an important isolating mechanism. We can perhaps glean some intuition from *E. editha,* in which oviposition behavior is clearly divided into pre-alighting and post-alighting phases (Parmesan *et al*. 1995). The pre-alighting phase is primarily visual (Parmesan et al. 1995), resulting in alighting biases towards or away from plants of particular growth form. These alighting biases are not learnt by *E. editha*: there was no difference in search efficiency between experienced female adults and recently eclosed naïve females. Insects from populations adapted to different hosts showed different alighting biases when released at the same site (Singer 2004).

How might local adaptation of visual alighting bias generate isolation in *E. aurinia?* The physical structure and visual appearance of *Lonicera* and *Succisa* could hardly be more different: *Lonicera* is a thin-stemmed shrub 1-4m tall with grey-green leaves 1-2 cm long,

while *Succisa* is a rosette 3-20 cm tall with red-green leaves 10-20 cm long close to the ground. If *E. aurinia*, like *E. editha*, has fixed visual alighting biases when searching for hosts, it is quite possible that a female from a *Succisa*-feeding population immigrating into a *Lonicera* site may be unable to find the local host, whereupon the post-alighting preferences become irrelevant. Since *E. aurinia* does not mate in association with hosts, visual responses to hosts should have no effect on the ability of immigrant males to find mates, although the female offspring of such males might suffer inefficient host search.

3.4.4 Differences Between the Genetic Associations in Allopatry and in Sympatry

The evidence presented above indicates that *E. aurinia* broadly lack the traits that are generally thought to be responsible for host-associated genetic differentiation under sympatric speciation. This is consistent with the lack of genetic differentiation between larvae found on the two hosts at LA. It is however not consistent with the level of host-associated genetic differentiation among populations found in the AMOVA and the Mantel correspondence tests. This differentiation suggests that there must be at least one mechanism in which host association generates isolation between populations on different hosts. A candidate for such an effect could be heritable pre-alighting host search, provided that migration among sites were mostly undertaken by females, as is the case in the related butterfly *Melitaea cinxia* (Kuussaari et al. 2000).

Whatever the trait responsible for host-associated differentiation among populations, the effect evidently breaks down in sympatry because no host associated genetic differentiation was detected at the dual-host use population LA. This should not surprise us. In these very sedentary butterflies, movement of individuals among sites must be several orders of magnitude less than movement within them. So, even if among-host movement always resulted in reduced gene flow compared to within-host movement, it is still possible that among-host movement between sites would be genetically negligible while among-host movement within-site would be sufficient to homogenize insects on the two hosts.

Table 3.1: Populations included in the study.

Population Code	Population Name	Coordinates		Sample Size	Host	
		Lat.	Long.		Lonicera	Succisa
CJ	Can Jorda	42.1443	2.50384	9		1
COL	Col Estelales	41.6635	1.97906	2	1	
COU	Coustouges	42.3584	2.64733	9		1
DAR	Darnius	42.3704	2.81783	12	1	
ELG	El Guix	41.8159	1.90441	4	1	
FORN	Col del Forn	42.5681	2.46861	10	1	
LA	LaBaraca	42.0457	2.62543	18	1	1
LNB	La Nou de Bergeuda	42.167	1.88627	2		1
MC	Mas Calc	41.9112	3.07234	14	1	
RED	Col de la Redoulade	42.9122	2.5145	3		1
TOR	Tordera	41.7308	2.7475	5	1	

Table 3.2: Population pairwise Φ_{ST} matrix (lower triangular matrix) and geographical distance matrix in km (upper triangular matrix).

	CJ	COL	COU	DAR	ELG	FORN	LA	LNB	MC	RED	TOR
CJ		68.90	26.57	36.08	61.62	47.16	14.87	51.10	53.72	85.29	50.17
COL	0.090		94.97	104.83	18.03	108.32	68.42	56.45	94.94	145.55	64.39
COU	0.103	0.060		14.11	86.07	27.54	34.78	66.29	60.84	62.47	70.20
DAR	0.223	0.071	0.163		97.47	36.15	39.41	80.10	55.17	65.12	71.27
ELG	0.196	0.000	0.148	0.180		95.66	65.01	39.02	97.53	131.72	70.72
FORN	0.218	0.080	0.124	0.021	0.173		59.45	65.46	88.35	38.40	95.81
LA	0.182	0.000	0.139	0.163	0.050	0.111		62.60	39.93	96.67	36.41
LNB	0.087	0.342	0.069	0.338	0.352	0.363	0.257		102.22	97.54	86.28
MC	0.251	0.163	0.210	0.051	0.250	0.091	0.201	0.367		120.29	33.61
RED	0.222	0.299	0.208	0.400	0.365	0.408	0.213	0.290	0.429		132.61
TOR	0.267	0.000	0.195	0.195	0.088	0.160	0.021	0.391	0.249	0.339	

Table 3.3: Host association matrix using Jaccard's coefficient (lower triangular matrix) and Sorensen's coefficient (upper triangular matrix).

	CJ	COL	COU	DAR	ELG	FORN	LA	LNB	MC	RED	TOR
CJ		1.00	0.00	1.00	1.00	1.00	0.33	0.00	1.00	0.00	1.00
COL	1.00		1.00	0.00	0.00	0.00	0.33	1.00	0.00	1.00	0.00
COU	0.00	1.00		1.00	1.00	1.00	0.33	0.00	1.00	0.00	1.00
DAR	1.00	0.00	1.00		0.00	0.00	0.33	1.00	0.00	1.00	0.00
ELG	1.00	0.00	1.00	0.00		0.00	0.33	1.00	0.00	1.00	0.00
FORN	1.00	0.00	1.00	0.00	0.00		0.33	1.00	0.00	1.00	0.00
LA	0.50	0.50	0.50	0.50	0.50	0.50		0.33	0.33	0.33	0.33
LNB	0.00	1.00	0.00	1.00	1.00	1.00	0.50		1.00	0.00	1.00
MC	1.00	0.00	1.00	0.00	0.00	0.00	0.50	1.00		1.00	0.00
RED	0.00	1.00	0.00	1.00	1.00	1.00	0.50	0.00	1.00		1.00
TOR	1.00	0.00	1.00	0.00	0.00	0.00	0.50	1.00	0.00	1.00	

Table 3.4: First ten eigenvalues from NMDS analysis of Nei's corrected average pairwise differences for (A) all populations, (B) population LA.

(A)

Dimension Number	Eigen-value	Individual Percent	Cumulative Percent	Cumulative Bar Chart
1	1101.01	48.79	48.79	□□□□□□□□□□□□□□□□□□□□□□□□□
2	642.56	28.47	77.26	□□□□□□□□□□□□□□□
3	88.99	3.94	81.21	□□□
4	79.15	3.51	84.71	□□□
5	5.5	0.24	84.96	□
6	0	0	84.96	□
7	-11.89	0.53	85.49	□
8	-24.12	1.07	86.55	□□
9	-61.8	2.74	89.29	□□
10	-100.74	4.46	93.76	□□□

(B)

Dimension Number	Eigen-value	Individual Percent	Cumulative Percent	Cumulative Bar Chart
1	6967.00	30.74	30.74	□□□□□□□□□□□□□□□□
2	4755.62	20.99	51.73	□□□□□□□□□□□
3	2530.48	11.17	62.89	□□□□□□□
4	1685.53	7.44	70.33	□□□□□
5	1469.40	6.48	76.82	□□□□
6	1030.00	4.55	81.36	□□□
7	857.72	3.78	85.15	□□□
8	681.35	3.01	88.15	□□□
9	627.50	2.77	90.92	□□
10	560.39	2.47	93.39	□□

Table 3.5: AMOVA analysis (A) between populations on *Succisa* and on *Lonicera*, (B) between individuals on *Succisa* and on *Lonicera* in the dual host-use population LA.

(A)

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among Groups	1	204.489	4.485	12.08
Among Populations Within Groups	8	456.979	4.446	11.98
Within Populations	60	1691.118	28.185	75.94
Total	69	2352.586	37.116	
$F_{CT} = 0.12$ ($P < 0.001$)				

(B)

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Between larvae on *Succisa* and *Lonicera*	1	33	1	4
Within each larval group	16	384	24	96
Total	17	417	25	
$F_{ST} = 0.04$ ($P = 0.11$)				

Table 3.6: Survey of territorial males defending potential habitat patches at DAR.

	With Lonicera	Without Lonicera
With territorial male	4	15
Without territorial male	16	58

Figure 3.1: Study sites.

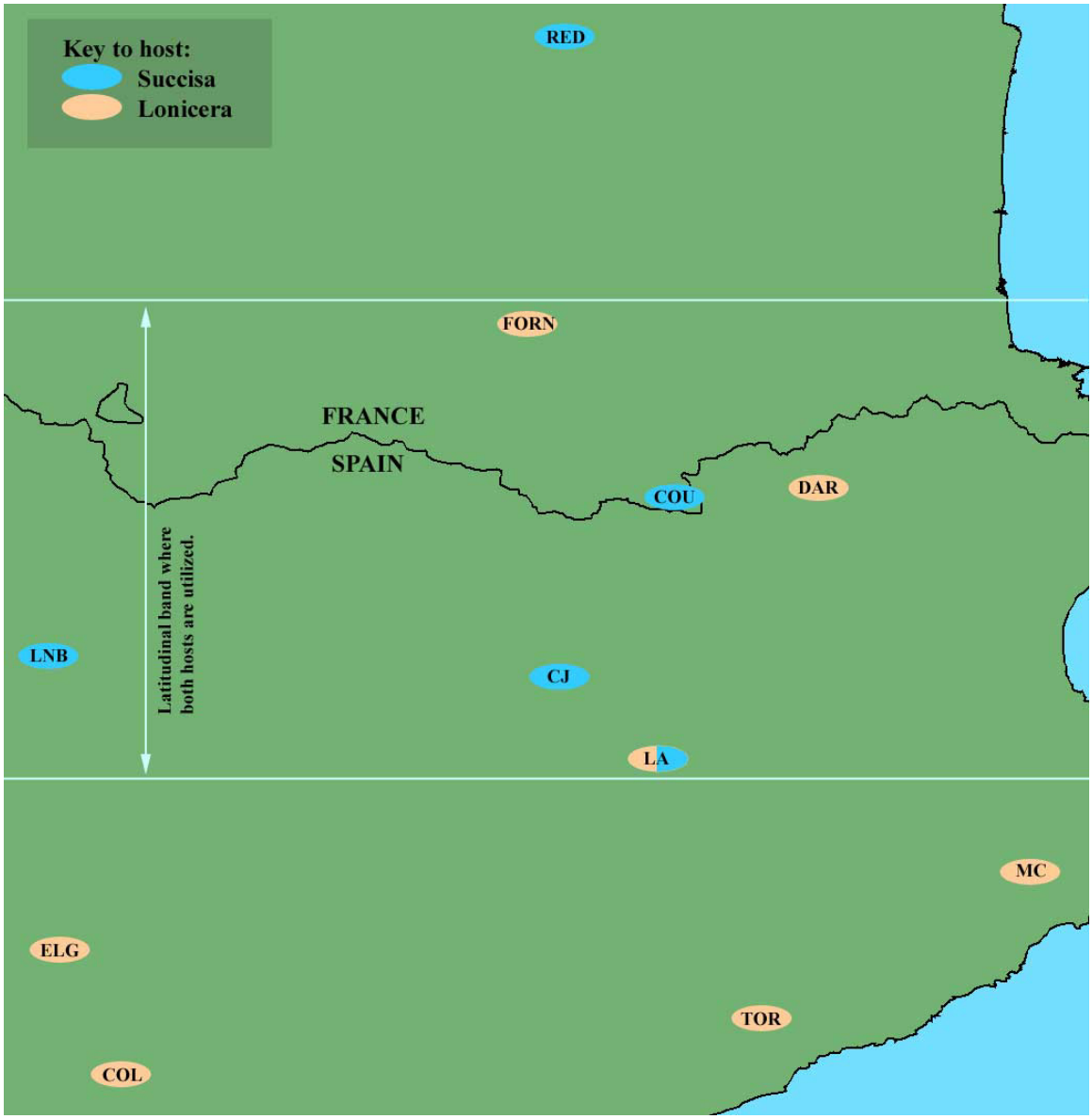


Figure 3.2: NMDS solution for all populations.

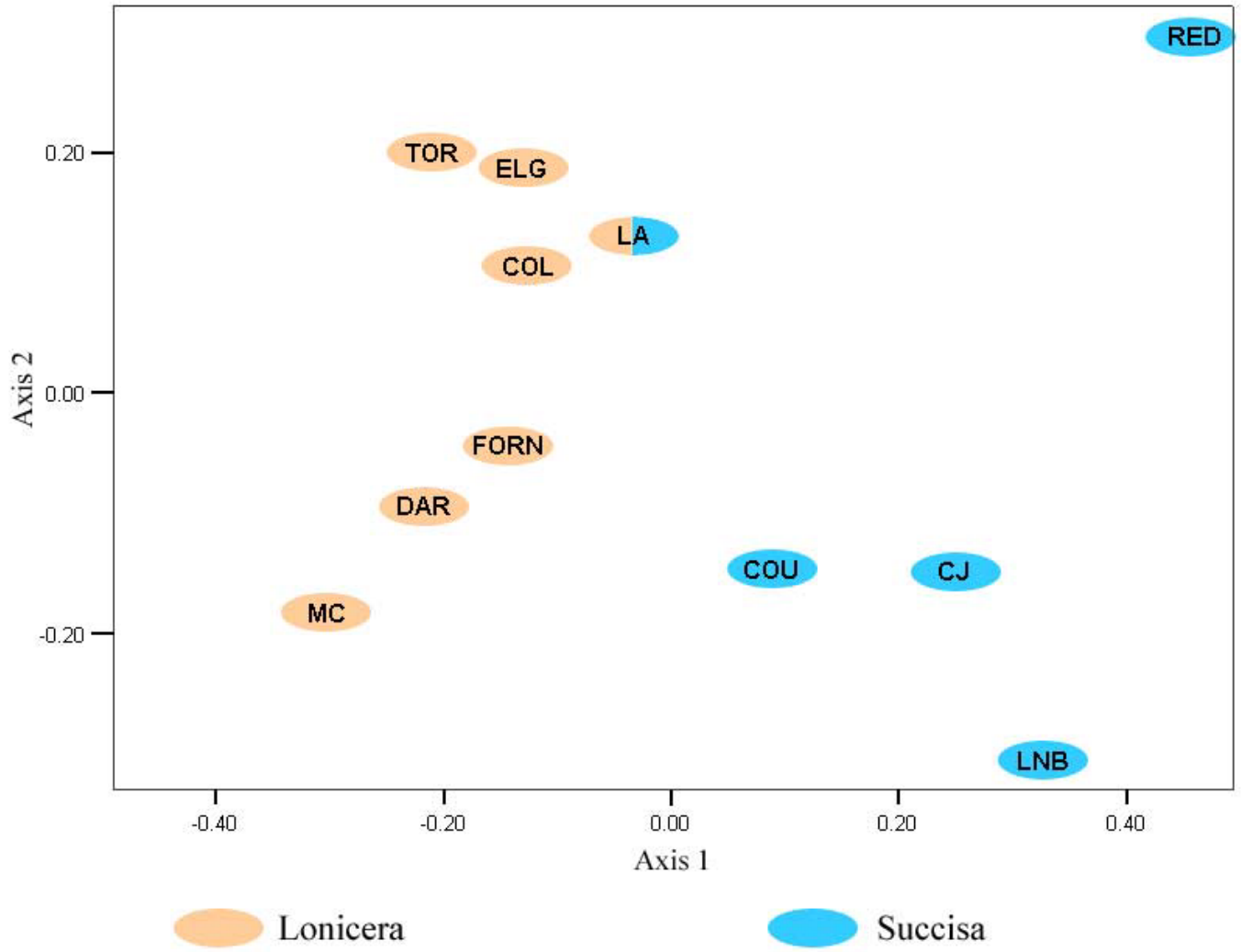


Figure 3.3: NMDS solution for the population LA.

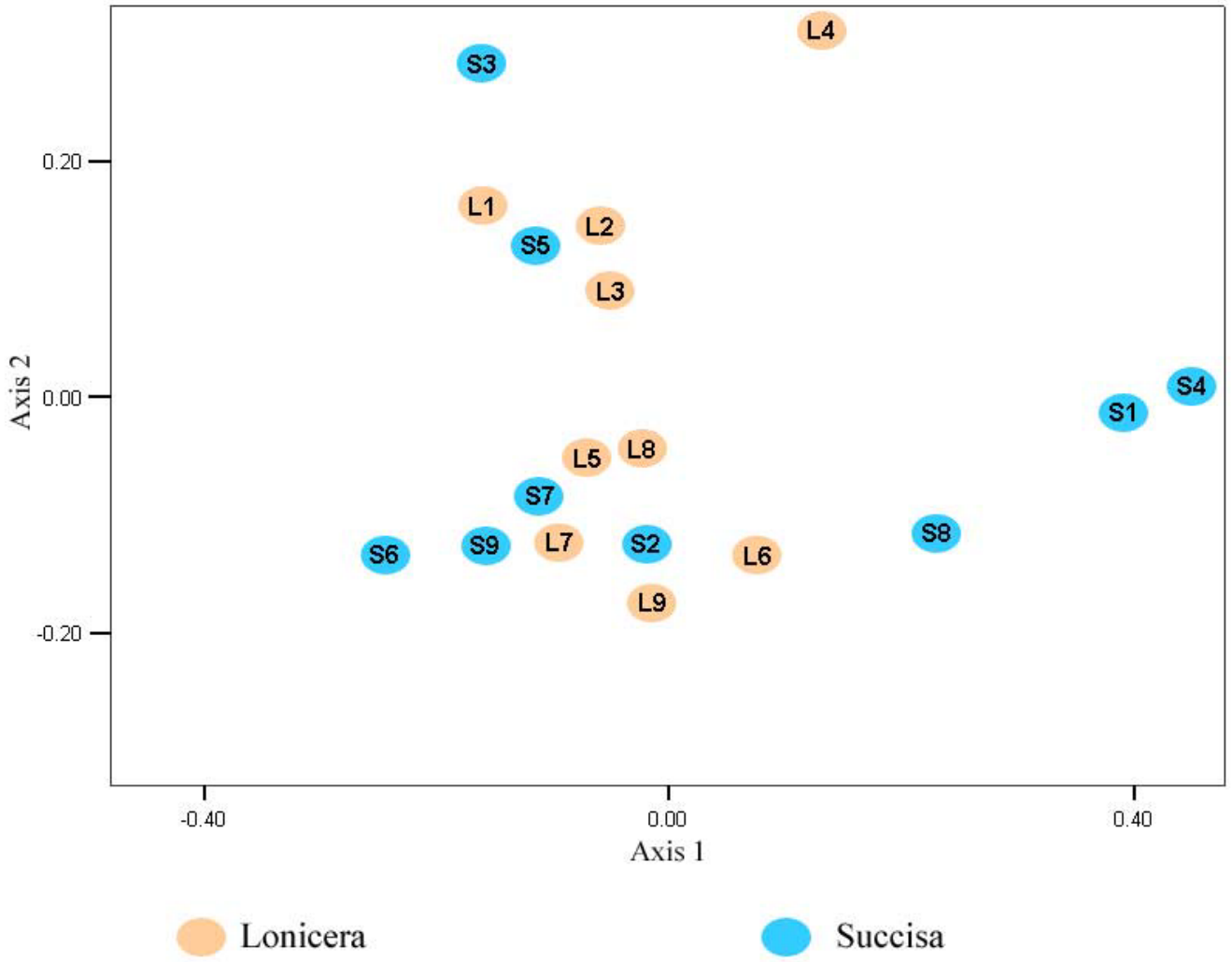
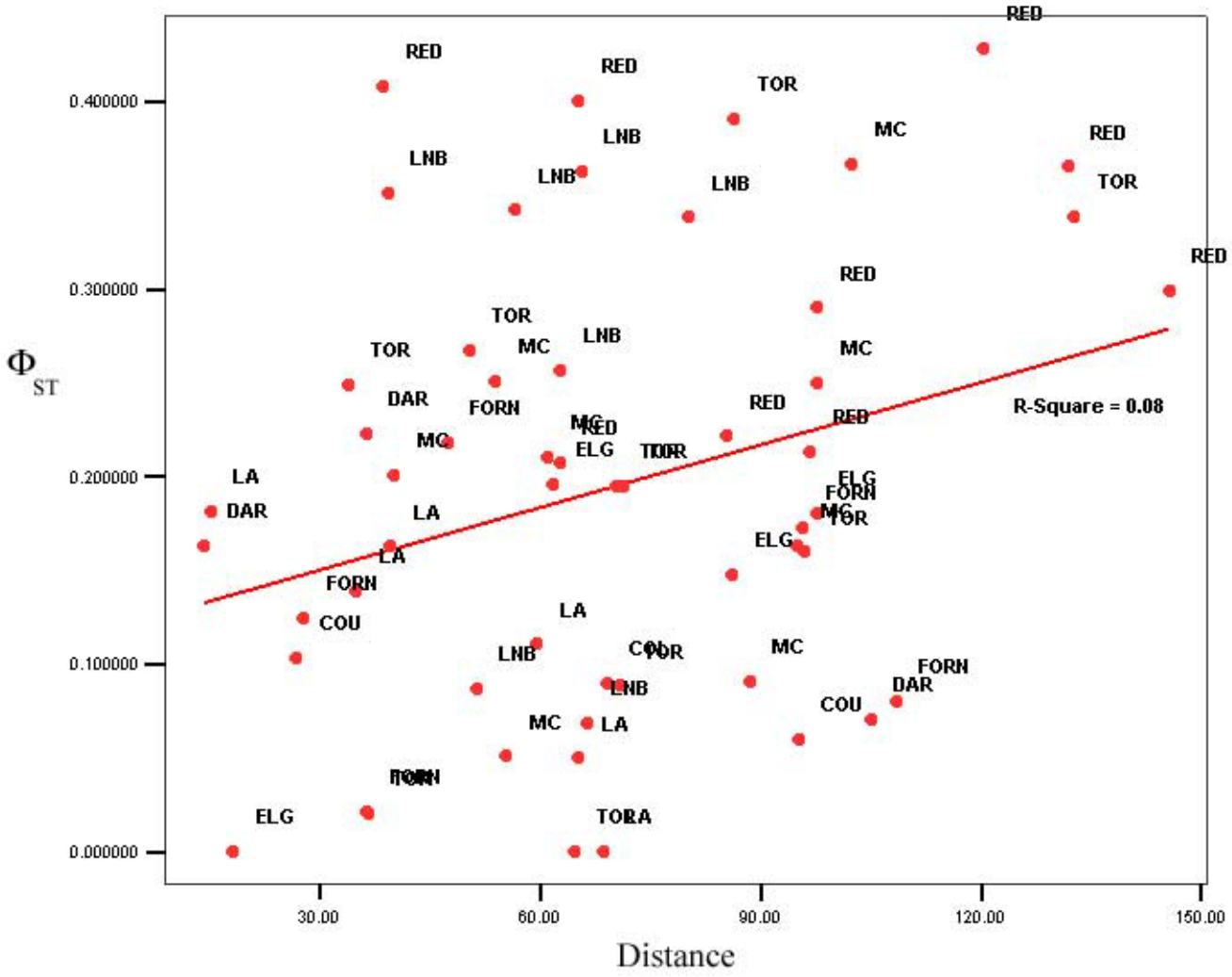


Figure 3.4: IBD plot of pairwise geographical distances versus pairwise Φ_{ST} values.



References

Abrahamson, W. G., M. D. Eubanks, et al. (2001). "Gall flies, inquilines, and goldenrods: A model for host-race formation and sympatric speciation." American Zoologist **41**(4): 928-938.

Anderson, P. R. and J. G. Oakeshott (1984). "Parallel geographical patterns of allozyme variation in two sibling *Drosophila* species." Nature **308**: 729-731.

Anthes, N., T. Fartmann, et al. (2003). "Combining larval habitat quality and metapopulation structure – the key for successful management of pre-alpine *Euphydryas aurinia* colonies." Journal of Insect Conservation **7**(3): 175-185.

Asher, J., M. Warren, et al. (2001). The Millennium Atlas of Butterflies in Britain and Ireland, Oxford University Press.

Begun, D. J. and C. F. Aquadro (1994). "Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation." Genetics **136**: 155-171.

Berlocher, S. H. and J. L. Feder (2002). "Sympatric Speciation in Phytophagous Insects: Moving Beyond Controversy?" Annual Review of Entomology **47**: 773-815.

Bonnet, E. and Y. Van de Peer (2002). "zt: a software tool for simple and partial Mantel tests." Journal of Statistical software **7**(10): 1-12.

Boughton, D. A. (1999). "Empirical evidence for complex source-sink dynamics with alternative states in a butterfly metapopulation." Ecology **80**(8): 2727-2739.

Bush, G. L. (1969). "Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis (Diptera: Tephritidae)*." Evolution **23**: 237-251.

Bush, G. L. and J. J. Smith (1998). "The genetics and ecology of sympatric speciation: a case study." Res. Popul. Ecol. **40**: 175-187.

Caillaud, C. M. and S. Via (2000). "Specialised feeding behaviour influences both ecological specialisation and assortative mating in sympatric host races of pea aphids." American Naturalist **156**: 606-621.

Craig, T. P., J. D. Horner, et al. (1997). "Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation." Evolution **51**(5): 1552-1560.

Craig, T. P., J. D. Horner, et al. (2001). "Genetics, experience, and host-plant preference in *Eurosta solidaginis*: implications for host shifts and speciation." Evolution **55**(4): 773-782.

Craig, T. P., J. K. Itami, et al. (2000). "The influence of host plant variation and intraspecific competition on oviposition preference and offspring performance in the host races of *Eurosta solidaginis*." Ecological Entomology **25**: 1-12.

Del Campo, M. L., S. Via, et al. (2003). "Recognition of host-specific chemical stimulants in two sympatric host races of the pea aphid *Acyrthosiphon pisum.*" Ecological Entomology **28**(4): 405-412.

Dres, M. and J. Mallet (2001). "Host races in plant-feeding insects and their importance in sympatric speciation." Phil. Trans. R. Soc. Lond. B **357**: 471-492.

Excoffier, L. (2001). Analysis of Population Subdivision. Handbook of Statistical Genetics. D. J. Balding, M. Bishop and C. Cannings, John Wiley & Sonds, Ltd**:** 271-307.

Excoffier, L., P. E. Smouse, et al. (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: applicable to human mitochondrial DNA restriction data." Genetics **131**: 479-491.

Feder, J. L. (1998). The apple maggot fly, *Rhagoletis pomonella:* flies in the face of conventional wisdom about speciation? Endless Forms, Species and Speciation. D. J. Howard and S. H. Berlocher. New York, Oxford, Oxford University Press**:** 130-144.

Feder, J. L. and K. E. Filchak (1999). "It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella,* and its implications for fitness trade-offs in phytophagous insects." Entomologia Experimentalis et Applicata **91**: 211-225.

Feder, J. L., T. Hunt, et al. (1993). "The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella.*" Entomologia Experimentalis et Applicata **69**: 117-135.

Feder, J. L., S. Opp, et al. (1994). "Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly." Proc Natl Acad Sci USA **91**: 7990-7994.

Feehan, J. and G. O'Donovan (1996). The Bogs of Ireland. Dublin, Ireland, University College Dublin.

Filchak, K. E., J. B. Roethele, et al. (2000). "Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*." Nature **407**: 739-742.

Funk, D. J., K. E. Filchak, et al. (2002). "Herbivorous Insects: Model Systems for the Comparative Study of Speciation Ecology." Genetica **116**: 251-267.

Hedrick, P. (1999). "Perspective: Highly variable loci and their interpretation in evolution and conservation." Evolution **53**(2): 313-318.

Hood, G. M. (2003). PopTools., http://www.cse.csiro.au/poptools.

Itami, J. K., T. P. Craig, et al. (1997). Factors affecting gene flow between the host races of *Eurosta solidaginis*. Genetic structure and local adaptation in natural insect populations: Effects of ecology, life history, and behavior. S. Mopper and S. Y. Strauss. New York, Chapman & Hall**:** 375-407.

Joyce, D. A. and A. S. Pullin (2001). "Phylogeography of the Marsh Fritillary *Euphydryas aurinia (Lepidoptera: Nymphalidae)* in the UK." Biological Journal of the Linnean Society **72**: 129-141.

Kuussaari, M., M. Singer, et al. (2000). "Local specialization and landscape-level influence on host use in an herbivorous insect." Ecology **81**(8): 2177-2187.

Luetgert, J. Range: Distance/Azimuth Program, U.S. Geological Survey; http://quake.wr.usgs.gov/~luetgert/Macintosh_Software/.

Mazel, R. (1986). "Structure et evolution du peuplement d'*Euphydryas aurinia* Rottemburg (Lepidoptera) dans le sud-ouest europeen." Vie Milieu **36**: 205-225.

Parmesan, C., M. C. Singer, et al. (1995). "Absence of adaptive learning from the oviposition foraging behaviour of a checkerspot butterfly." Animal Behaviour **50**: 161-175.

PE-Biosystems Genotyper.

Saccheri, I., I. Wilson, et al. (1999). "Inbreeding of bottlenecked butterfly populations: Estimation using the likelihood of changes in marker allele frequencies." Genetics **151**: 1053-1063.

Schneider, S., D. Roessli, et al. (2000). Arlequin: A software for population genetics data analysis., Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Singer, M. C. (2004). Measurement, Correlates, and Importance of Oviposition Preference in the Life of Cherkerspots. On the Wings of Checkerspots. P. Ehrlich and I. Hanski. Oxford, NY, Oxford University Press.

Singer, M. C., C. Stefanescu, et al. (2002). "When random sampling does not work: standard design falsely indicates maladaptive host preferences in a butterfly." Ecology Letters, **5**: 1-6.

Thomas, Y., M. Bethenod, et al. (2003). "Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis hubner*. I. Sex pheromone, moth emergence timing, and parasitism." Evolution **57**(2): 261-273.

Via, S. (1991). "The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones." Evolution **45**: 827-852.

Via, S. (1991). "Specialised host plant performance is not altered by experience." Ecology **72**: 1420-1427.

Via, S. (1999). "Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice." Evolution **53**(5): 1446-1457.

Via, S., A. C. Bouck, et al. (2000). "Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environment." Evolution **54**(5): 1626-1637.

Wahlberg, N., T. Klemetti, et al. (2002). "Dynamic populations in a dynamic landscape: The metapopulation structure of the marsh fritillary butterfly." Ecography **25**(2): 224-232.

Wahlberg, N., T. Klemetti, et al. (2002). "Metapopulation structure and movements in five species of checkerspot butterflies." Oecologia **130**: 33-43.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**: 1358 - 1370.

Wood, T. K. and M. C. Keese (1990). "Host plant induced assortative mating in *Enchenopa* treehoppers." Evolution **44**: 619-628.

Wright, S. (1978). Evolution and the Genetics of Populations. Chicago, University of Chicago Press.

Chapter 4: Effect of Landscape Features on the Effective Isolation of *Euphydryas editha* Populations

Abstract. "Isolation by distance," the reduction of gene flow with increasing geographic separation, can cause genetic differentiation to increase with geographic distance. For organisms with patchy spatial distributions, this effect can be augmented by the influence of the matrix of unsuitable landscape that separates suitable patches. The objective of this study is to investigate whether a Geographical Information System can be used to model the influence of the matrix on effective isolation between populations of Edith's checkerspot butterfly, *Euphydryas editha*. Eight populations each from Sequoia National Park / National Forest and Yosemite National Park were sampled and assayed using Amplified Fragment Length Polymorphisms (AFLP). ArcGIS was used to perform cost-surface modeling to test three metrics of effective isolation: accumulated costs, topographical distances along least cost paths, and least cost path lengths. These metrics were generated using three ArcGIS cost models that make different assumptions about how landscape features contribute to effective isolation. The less permeable matrix type, forest, was assigned different relative resistance values for sensitivity analysis purposes. Mantel correspondence tests were then performed to assess the correlation between the different metrics and genetic differentiation. Genetic differentiation within both Sequoia and Yosemite was found to be small but statistically significant ($\Phi_{ST} = 0.03$, $P < 0.001$). At Sequoia, genetic differentiation was not correlated with either straight line distances or elevation. None of the correlation tests between genetic differentiation and effective spatial distances for Sequoia were statistically significant. At Yosemite, strong IBD was detected when straight line distances were used ($r = 0.66$, $P < 0.001$). Effects of elevational differences on genetic differentiation could not be detected. Of the three effective distance metrics tested, the accumulative cost metric produced similar correlations over all friction values for all cost models. Accumulative cost metrics computed with the basic cost model using a resistance value of 11.5 resulted in an

increased correlation between genetic differentiation and spatial distances ($r = 0.77$, $P < 0.001$).

4.1 INTRODUCTION

4.1.1 Background

Levins (1969) introduced the concept of the metapopulation as an interconnected set of subpopulations that together function as a demographic unit. The concept grew out of a need to address the simplifying assumptions (Simberloff 1974) of island biogeography, first developed by MacArthur and Wilson (1963; 1967). The Levins model was not without its own limitations: one drawback was that it was spatially implicit, and assumed that organisms can easily locate unoccupied patches regardless of distances between patches and regardless of the quality of intervening landscape matrix (Turner et al. 2001). Since then, metapopulation models have grown in complexity to include effects of habitat connectivity, habitat fragmentation, corridor characteristics, patch quality, and patch permeability, among other factors (Hanski and Gaggiotti 2004).

The role of matrix quality. The presence of intervening "matrix" between habitat patches may prevent organisms from leaving their natal patches and may form population sinks when those that do leave perform poorly in intervening habitat. However, there is still very limited knowledge of the factors affecting movements between patches within a habitat patch network (Moilanen and Hanski 1998; Ricketts 2001). None of the three indices commonly used to quantify landscape connectivity – the gamma index, the alpha index, and the connectivity matrix – take into account the quality of unsuitable habitat connecting habitat patches. Yet, Henein and Merriam (1990) showed in a simulation study just how important matrix quality can be: when a habitat patch with a low quality dispersal route is added to an existing habitat network, it can act as a population sink and reduce the equilibrium number of individuals in the metapopulation (Fahrig and Merriam 1994). Fahrig and Merriam (1994) list other examples where low quality dispersal routes

acted as migration barriers. Fahrig (2001) also demonstrated in a simulation study how total habitat abundance could be traded off for a higher quality intervening matrix.

If there is a dearth of studies that address how movements are affected by the intervening matrix, there is only one published study (Keyghobadi et al. 1999) that addresses how the matrix affects genetic diversity within the patch network. Studies that investigate population genetic structure in fragmented landscapes typically take landscape characteristics into account only after the population genetic analysis as a means to explain biogeographic patterns: there are very few studies whose primary aim is to correlate genetic variation to the type of intervening landscape matrix.

Effect of intervening landscape matrix on habitat isolation. Gene flow is likewise reduced when the matrix reduces movement between patches (Peterson and Denno 1998). Therefore, the nature and extent of the matrix should influence genetic differentiation among populations. Roland et al. (2000) and Keyghobadi et al. (1999) found strong effects of distance through forest on genetic distances among populations of *Parnassius smintheus* butterflies, which moved more readily through open meadow than through forests. Forests were found to be twice as resistant to butterfly movement as open meadows. Differential rates of movement through different habitat types were also found in a mark-release-recapture study of 21 butterfly species in the Rocky Mountains (Ricketts 2001). In that study, willow and conifer offered different resistance to butterfly movement between habitat patches, modifying the effective patch isolation between suitable habitats. Norberg *et. al.* (2002) compared the movement rates of three *Melitaeini* species to three other species in different subfamilies of the Nymphalidae (*Brenthis ino*, *Aphantopus hyperantus*, and *Clossiana euphrosyne*). They showed that the *Melitaeini* butterflies moved at considerably lower rates than the other species through shady habitat.

The subjects of the present study, *Euphydryas editha* butterflies, are also *Melitaeini*. Like the *Melitaeini* studied by Norberg *et al.* (2002), they are sun-loving insects that are almost invariably seen in open patches of habitat with direct insolation. Like many

butterflies (Watt 1969) they typically fly at body temperatures of 35 – 42^{o}C and are unable to fly at much lower body temperatures (Parmesan, C., personal communications). In consequence, flight through shady habitats in cool air would risk enforced alighting. Flight over the treetops would be energetically expensive but carry less risk of immobility.

While Ricketts (2001) found that the nature of the matrix between habitat patches was important, Moilanen and Hanski (1998) concluded that it was not. In the latter study, the incorporation of information about matrix quality failed to significantly improve the fit of a model predicting patch occupancy for the Glanville fritillary butterfly *Melitaea cinxia*. A GIS was used to modify the patch isolation between suitable habitats by increasing interpatch distances for "difficult" intervening matrix and decreasing distances for "easy" matrix.

Effect of elevation changes on habitat isolation. Roland et al. (2000) found that the recapture rates of *Parnassius smintheus* butterflies was negatively correlated with changes in elevation. This effect was still significant after controlling for distance through forests and meadows. They suggest that these elevation biases cause valleys to act as barriers to movement in their study organism. Weiss and Weiss (1998) found an effect of elevation on the phenology of emerging *E. editha bayensis* adults: microclimate differences over a 300 m elevation gradient were sufficient to cause a difference of one week in mean adult eclosion. Similarly, Peterson (1997) found an effect of elevation on adult oviposition: a mark-release-recapture study of the lycaenid butterfly *Euphilotes enoptes* found that female insects showed a preferential movement uphill in response to senescing low-elevation host-plant patches. In contrast, Keyghobadi et al. (1999) did not find a significant effect of elevation change on dispersal.

Objective of this study. The objective of this study is to investigate whether a Geographical Information System can be used to model the effective isolation between habitats. One method to test for the effects of isolation by distance (IBD) (Wright 1943)

is to determine whether genetic differentiation is correlated with straight line distances between populations. However, such straight line measures are not necessarily good estimates of the paths that organisms take to circumnavigate unsuitable landscape matrix. Straight line measures also do not take into account the actual distance traversed along a topographical surface, henceforth referred to as topographical distance. In landscapes where there is substantial variation in topography we expect the effects of the discrepancy between straight line distances and topographical distances to increase as distances increase. This discrepancy can adversely affect any attempt to estimate geographical isolation from straight-line distances. This problem is compounded by the influence of differences in matrix type.

This study seeks a methodology to overcome these shortcomings in estimating effective isolation. It does so by testing three alternate effective geographical distance metrics (henceforth referred to as "effective distance metrics") that model isolation between populations separated by heterogeneous landscape matrix. These metrics are used in correlation tests against genetic differentiation to see which model best explains genetic variation between populations that is attributable to spatial separation.

4.1.2 Study system.

The butterfly *Euphydryas editha* occurs in diverse habitat types, including chaparral, oak woodland, serpentine grassland, montane coniferous forest and high elevation tundra. *E. editha* uses a diversity of host plants in the families *Antirrhinaceae* and *Orobanchaceae*. Populations in the Sierra Nevada utilize five different host genera for oviposition. 70% of these populations are monophagous, while 30% use from 2 – 4 host genera. Monophagous populations on one host genus often occur interdigitated with populations on a different host. Population structure is also variable within this species: those found in Sequoia National Forest and Yosemite National Park typically exist as large metapopulations, while in some regions most of the insects exist in well-defined and relatively isolated populations.

4.2 MATERIALS AND METHODS

4.2.1 Sampling

Specimens were collected from eight populations within Sequoia National Park / National Forest, and eight populations within Yosemite National Park. Movement patterns of the butterflies among habitat patches have been extensively studied by mark-recapture at Sequoia by Thomas & Singer (1987) and Boughton (2000), but only a small study has been conducted at Yosemite (Gilbert L.E., Singer M.C., unpublished). A total number of 106 specimens were collected for Sequoia, for an average of 13 specimens per population. The respective numbers for Yosemite are 90 and 11 respectively. The populations examined in this study are listed in Table 4.1, and their geographical relationships are depicted in Figure 4.1.

Both adult insects and larvae were used for the molecular assay. Larvae were collected from the field between 1999 and 2003. In populations where the adults were known to oviposit on more than one host species, an attempt was made to collect an equal number of larval specimens from each host species. These specimens were labeled with the host plants on which they had been collected and frozen at -80 °C. Freshly dead or dying adults were similarly preserved. In order to minimize the impact of sampling on certain populations, adults were caught and released after a small wing clipping was taken. The clippings were preserved in 100% EtOH for the molecular assay.

4.2.2 AFLP Assay

Sample preparation. DNA was isolated from specimens either frozen at -80 °C or preserved in 100% EtOH. For adult specimens, a single leg was removed for DNA isolation, and the rest of the body stored for future use. In order to minimize possible contamination from parasitoids in the larval body, DNA was extracted from the heads of

larvae, which are least likely to contain parasitoids. Wing clippings were completely used in the extraction process.

DNA extraction. Qiagen DNeasy Tissue kits (Catalog No. 69506) were used for DNA extraction. The Qiagen protocol was followed except for the following four modifications: (1) A CTAB extraction buffer (10% 1 M Tris-HCL pH 8.0, 28% 5 M NaCl, 4% 0.5 M EDTA pH 8.0, 2% CTAB) was used instead of Qiagen's tissue buffer. (2) Prior to extraction, 2 µl of beta-mercaptoethanol per 1ml of extraction buffer was added to the CTAB buffer. (3) Samples were crushed in a 1.5ml Eppendorf tube in liquid nitrogen, and then homogenized in 200 µl of the extraction buffer. (4) The DNA was eluted in two 100 µl washes of buffer AE for a final DNA volume of 200 µl.

AFLP PCR reaction. The Applied Biosystems (ABI) AFLP protocol for small plant genomes (www.appliedbiosystems.com; protocol 4303146) was used with the following six modifications: (1) 30 ng of DNA was used for restriction-ligation and incubated using a thermocycler with a heated lid for 3 hours. (2) 120 µl of $TE_{0.1}$ was added to the restriction-ligation product instead of the recommended 189 µl. (3) Pre-selective amplification with the regular genome preselective primer mix was next performed in 10 µl reactions. (4) Instead of the recommended 20 cycles of denature – anneal – extension, 25 cycles were used instead. (5) The pre-selective product was not diluted for the selective amplification. (6) For the selective PCR, the 25 cycles of denature – anneal – extension (20 s at 94 °C, 30 s at 56 °C, 2 min at 72 °C) were used instead of the recommended 20.

The three primer pairs from ABI used for this study were: (1) EcoRI –ACA / MseI – CAT, (2) EcoRI – ACA / MseI – CTG, and (3) EcoRI – AAG / MseI – CTG. 3 µl of the selective product was mixed with 0.4 µl of GeneScan 500 Rox and 6.6 µl of Hi-Di Formamide. Samples were processed on an automated ABI 3100 Genetic Analyzer.

4.2.3 Analysis

Scoring of AFLP data. Raw fragment data from the 3100 Genetic Analyzer were scored in Genotyper v3.6 (PE-Biosystems). A custom written Genotyper macro was used to export the scored data to a custom written PC program, AFLPal (see Chapter 2 "An efficient AFLP protocol for butterflies: from laboratory to analysis"). In most cases, a given locus for an individual is scored as either absent or present, depending on whether an AFLP fragment was generated for that locus. However, certain populations were characterized not only by the absence or presence of a fragment, but also by the number of copies of that fragment. This was reflected by the fragment peak height in the Genotyper software. Peak heights with bimodal frequency distribution reflect individuals from two different populations and should be scored as such (personal observation). AFLPal was used to analyze the frequency distribution of the number of fragments associated with any one locus. Based on that analysis, the program allows the user to control how the absence-presence matrix is to be generated. In addition, AFLPal allows the user to specify parameters to denote a fragment that is present (scored as '1'), absent (scored as '0'), or ambiguous (scored as '?'). The latter state can be used by many programs to denote missing data.

Genetic analysis. Arlequin (Schneider et al. 2000) was used to analyze the absence – presence matrix generated by AFLPal. The AMOVA (Excoffier et al. 1992) module in Arlequin was used to determine Φ statistics, equivalent to the Weir and Cockham's (1984) θ statistics (Excoffier 2001). All Φ statistics were computed for statistical significance with 5040 permutations. The population pairwise Φ_{ST} matrix calculated by Arlequin was used in full and partial Mantel correspondence analyses to calculate r, the Pearson correlation coefficient, using the program ZT (Bonnet and Van de Peer 2002). All significance tests were performed with 10000 permutations. To test for effects of IBD, the correspondence test was done between the Φ_{ST} matrix and a pairwise geographical distance matrix between populations. The program Range (Luetgert, USGS) was used to

calculate the geographical separation between populations measured as straight line distances between two points. Geographical coordinates were acquired from on-site readings using a hand-held GPS receiver or estimated from a topographical map. The elevation above sea level for each population was determined in a similar fashion. Differences in elevation may act as a barrier to gene flow, mediated either through behavior (such as a tendency to fly uphill versus downhill), physiology (such as operational constraints of flight muscles related to ambient temperature), or host phenology (such host senescence). To test whether such effects are present, Mantel correspondence tests were conducted between the Φ_{ST} matrix and a matrix representing differences in elevation between populations.

4.2.4 GIS Modeling

Overview. In order to model the effects of landscape features, the ArcGIS 8.1 (ESRI) suite of programs was used to model the effects of topography and vegetation cover of the Sequoia and Yosemite study sites. An overview of the process is given below, and the details are described later. To model an effective distance metric between populations A and B, the procedure is as follows:

(1) Classify image. A digitized image of the study area is first classified into open and forested areas.

(2) Generate friction surface. Pixels on the digitized image are assigned relative 'friction' values. Friction refers to the impediment to movement, or resistance, imposed by various landscape elements as an organism moves across a heterogeneous landscape. In this study, open areas like meadows and barren terrain are given a relative friction value of 1. Relative friction values for forested areas are assigned some non-negative, non-zero real number. For example, if the relative friction value for forested areas is ten, this implies that it is ten times more costly to

traverse through the forest than through meadows. The set of pixels with assigned relative friction values is known as a friction surface.

(3) Generate accumulated cost-surface. Given a friction surface and a point on the surface representing a population, a cost model is used to calculate the accumulated cost of emigrating from that population. Each population is uniquely associated with its own cost-surface. Cost models can be parameterized to model different resistances to dispersal posed by landscape elements such as topography and vegetation cover.

(4) Calculate the effective distance metric between populations. Given the accumulated cost-surface, say for population A, the accumulated cost expended in traveling to population B can be determined by examining the value of the pixel under the cell corresponding to population B. An accumulated cost is one example of an effective distance metric. Further modeling steps can be conducted on the accumulated cost-surface to model other types of effective distance metrics.

(5) Perform Mantel correspondence test. For a given effective distance metric, the effective distances between populations computed from the previous step are used to create a population pairwise effective distance matrix. A Mantel correspondence test between this matrix and the population pairwise Φ_{ST} estimates is performed to estimate the amount of correlation between genetic differentiation and the effective distance metric.

In this study, three effective distance metrics were generated using five friction values against three cost models. The effective distance metrics and cost models used in the modeling are described later. A Visual Basic for Applications (VBA) program was written within the ArcMap 8.2 environment to automatically iterate through the steps described above with different friction values and different cost models. A high level algorithm is given below. See Figure 4.3 for a graphical version.

#Define PARK = [Yosemite, Sequoia]

#Define FRICTION_VALUE = [6.9, 9.2, 11.5, 13.8, 16.1]

#Define COST_MODEL = [CostDistance, PathDistance(SecPower=1.0),
PathDistance(SecPower=1.5)]

For each *Park* in PARK

For each *FrictionValue* in FRICTION_VALUE

Generate *FrictionSurface* from classified image using *FrictionValue*

For each *CostModel* in COST_MODEL

For each *Population* in *Park*

Generate *AccumulatedCostSurface* by applying *CostModel* on *FrictionSurface* for *Population*

Calculate *EffectiveDistanceMetrics* traveling from *Population* to other populations in *Park*

Generate population pairwise *EffectiveDistanceMetrics* matrices

Perform Mantel test, Φ_{ST} matrix X *EffectDistanceMetrics* matrices

The term *EffectiveDistanceMetrics* is used to denote the set of three effective distance metrics used in the study. Details of each of the major steps are given below.

Step 1: Classify image. Eight meter resolution georeferenced digital orthophoto quarter quads (DOQQs) of the study sites purchased from www.mapmart.com were used to model the study sites in the GIS. DOQQs are digital photos that have been georeferenced and divided into quadrangles corresponding to USGS topographic map quadrangles. ER Mapper 6.3 was used to mosaic, color-balance, and classify the individual DOQQs. At the 8m resolution, it was evident from a visual inspection that the resultant map featured an abundance of forested areas which were classified as permeable due to gaps between trees. Experience with *E. editha* suggested that these gaps would in fact be effectively

impermeable to adult insect movement because of insufficient direct insolation (personal observation). In addition, preliminary tests with the *CostDistance* model (see below, "Step 3: Generate accumulated cost-surface") revealed that, at the 8m resolution, the resultant cost-surface was almost identical to a cost-surface created using a DOQQ of a barren landscape without vegetation. For the purpose of this analysis, the 8m resolution DOQQ was too fine-grained. In order to increase the granularity level, SpatialAnalyst in ArcMap 8.2 was used to average each pixel value based on a 10 x 10 window using the ArcGIS RasterCalculator function *focalmean*. The resultant image was then manually classified into open and forested areas by identifying the range of pixel values that corresponded to each type of area.

Step 2: Generate friction surface. From my own observations of *E. editha*, I assume that the most important categorization of matrix is bimodal, as forest or non-forest, with the forest being less permeable to the butterflies than more open habitat types. But how much less permeable? An initial guide can be found in the work of Ricketts (2001), who used a mark-recapture approach to estimate the relative resistance of different matrix types to the movement of 21 butterfly species. For two species of *Melitaeini* butterflies (*Phyciodes campestris* and *Chlosyne palla*), Ricketts estimated the relative resistance for movement through conifers relative to open meadows as 11.5. At the study sites in Sequoia, forested areas comprise primarily red fir, Jeffrey pine, and lodge pole pine. Forested areas at the Yosemite study sites comprise various fir trees and lodgepole pine. The relative resistance value for firs and pines was initially assumed to be identical to that found by Ricketts for "conifers." The DOQQ from the previous step was reclassified by assigning a value of 11.5 to forested areas, and 1 to pixels which corresponded to open areas (wet meadows, dwarf alpine shrub, barren areas). The resultant 'friction surface' was then used to create an accumulated cost-surface (hereby referred to as 'cost-surface') using one of three cost models (see below, "Step 3: Generate accumulated cost-surface"; see Figure 4.3). In order to perform sensitivity analysis, four additional friction surfaces

were created with friction values 6.9, 9.2, 13.8, and 16.1. These represent successive increments or decrements of 20% from 11.5.

Step 3: Generate accumulated cost-surface. An accumulated cost-surface was created for each population within Sequoia and Yosemite for each of the five friction values using three different cost models (see Figure 4.3). In ArcGIS, the generation of an accumulated cost surface starts at the cell (pixel) representing the origin of dispersal. A spreading algorithm (Douglas 1994; Collischonn and Pilar 2000) searches its eight neighboring cells, stopping at the first cell that does not have an accumulated cost assigned to it. The algorithm calculates the accumulated cost of reaching that cell, and then repeats the process until all cells have been assigned an accumulated cost. Figure 4.2 illustrates how accumulated costs are calculated using a terrain represented by just three cells (ESRI 2002).

Three cost models are tested in this study. These models generate accumulated cost-surfaces based on the principles described above. The models are described next.

Model: *CostDistance*. *Costdistance* is a basic cost modeling function which creates a cost-surface by accumulating cost values as it simulates an individual emigrating out into the surrounding area from a given source. The cost value is calculated by multiplying the dimension of a pixel, 8m in the case of 8m resolution DOQQs used in this study, by the corresponding value on the friction surface, which may be a 1 (for open areas) or one of the five friction values (for forested areas; see above).

Model: *Pathdistance(Secpower=1.0)*. *Pathdistance* is a more advanced cost-modeling function which can be parameterized to assign differential costs contingent on factors like slope angle. At Sequoia, parts of the intervening terrain between populations comprise Californian foothill chaparral. In summer, during the adult *E. editha* flight season, this terrain becomes too hot and dry to be considered a viable migration corridor. It was assumed that emigrating adults are unlikely to descend from the cooler upper elevations down to the lower elevations and then ascend to higher elevations again. To impose this

constraint, *Pathdistance* was parameterized to bias the cost-surface towards moving uphill. The magnitude of this bias, contingent on the slope angle, is shown in Figure 4.4A. This bias, called the vertical factor, decreases for increasingly positive angles (uphill movement), and increases for increasingly negative angles (downhill movement). The bias for the latter is controlled by the secant of the downhill angle. Imposing this constraint makes it is more costly to descend than to ascend, even if the descent ends up in a viable, upper elevation habitat patch. A more realistic alternative would be to impose a cost penalty for descent only if a butterfly goes below a certain elevation for a given period in the season, or, the cost penalty only becomes effective on a sustained descent path after a certain threshold distance. However, this is beyond the scope of the current study.

Model: *Pathdistance(Secpower=1.5).* *Pathdistance* was again used as the cost modeling function. However, a greater bias against downhill movement was imposed, as shown in Figure 4.4B. In this case, the secant of the downhill angle is raised to a power of 1.5. Hence, this model is referred to as the *Pathdistance(secpower=1.5)* model, whereas the former model is referred to as the *Pathdistance(secpower=1.0)* model.

Step 4: Calculate the effective distance metric between populations. Three distance metrics were computed for this step, described below.

Metric: Accumulated cost metric. The accumulated cost of traveling between two populations can easily be determined by examining the accumulated cost-surface for the 'source' population, say population A. The pixel value on that cost-surface that corresponds to population B, the 'destination' population, represents the accumulated cost of traveling from A to B.

Metric: Path length metric. For any two given populations A and B, the ArcGIS RasterCalculator function *CostPath* was used to solve for the least cost path from A to B using A's accumulated cost-surface. For the return direction B to A, B's accumulated cost-surface is used instead. The 'to' and 'from' least cost paths may be different (Figure

4.3). The path length between A and B is calculated as the average lengths of the paths from either direction.

Metric: Topographical distance metric. The least cost path solutions from the previous step were overlaid on a topographical distance surface to calculate the topographical distance along the least cost path (Figure 4.3). To create the topographical distance surface, a slope surface was first created from a digital elevation model (DEM) of the study area. With a known slope angle and a fixed cell size (8m), the topographical distance surface can be calculated using basic trigonometry. As slope angles approached 90^{o}, topographical distances approached infinity. To limit this asymptotic rise in topographical distances, slope angles of more than 60^{o} were assumed to be 60^{o}. On a 8m resolution DEM, this meant a maximum topographic distance of 16m when the slope angle was 60^{o} or more. The sum pixel values of this topographical distance surface under the least cost path gave the topographical distance between two points along the least cost path. Similar to the path length metric (above), the topographical distance between two points is calculated using the average values from both directions.

Step 5: Perform Mantel correspondence test. For a given effective distance metric calculated using a given friction value computed under a given cost model, a matrix of population pairwise effective distance metric is computed. A Mantel test of this matrix against a population pairwise Φ_{ST} matrix is then performed to obtain a Pearson correlation coefficient. For each of the cost models tested, fifteen correlation coefficients were generated: five friction values for each of the three metrics. All fifteen correlation coefficients were plotted on the same graph for each cost model. Figure 4.7A shows the correlation coefficient plot for a hypothetical cost model. Assuming that the metrics accurately reflect the effective isolation posed by landscape elements, the correlation values should ideally co-vary across all friction values for all metrics. That is, for any two discrete consecutive friction values, the correlations should either increase or decrease in tandem. The correlations produced should also ideally exhibit little or no variation over a moderate range of different friction values.

4.3 RESULTS

4.3.1 AFLP Analysis

A total of 547 AFLP loci were identified, 546 which were polymorphic. The primer pairs EcoRI –ACA / MseI – CAT, EcoRI – ACA / MseI – CTG, and EcoRI – AAG / MseI – CTG generated 255, 132, and 160 markers respectively. Table 4.2 shows the pairwise Φ_{ST} matrix generated by Arlequin for both Sequoia and Yosemite. One Φ_{ST} value in the Yosemite matrix was negative at -0.001. This Φ_{ST} value was converted to zero for further analysis. Negative values can sometimes occur in the absence of genetic structure when the true value of the fixation index is zero (Excoffier 2004).

The eight populations within Sequoia demonstrated a low degree of genetic differentiation. Φ_{ST} was estimated at 0.03 ($P < 0.001$). The minimum, maximum, and mean distances between populations are 0.8km, 15.8km, and 8.7km respectively. The mean elevation between populations is 115m. No significant relationship was found between genetic differentiation and geographical distance or elevational distance (Table 4.3, Fig 4.4A).

Similarly, the eight populations within Yosemite demonstrated a low degree of genetic differentiation. Φ_{ST} was estimated at 0.03 ($P < 0.001$). The minimum, maximum, and mean distances between populations are 1.7km, 17.2km, and 9.3km respectively. The mean elevational distance is 367m. There is no evidence that genetic structuring is correlated with elevational distances. However, there is a significant relationship between geographical separation and genetic differentiation ($r = 0.66$, $P < 0.001$) (Table 4.3, Fig 4.4B). A previous mark-recapture study (Singer MC, Gilbert L., unpublished) found that adult insects at Yosemite are more sedentary than insects at Sequoia. This may help explain the why IBD was found at Yosemite, but not at Sequoia.

At Yosemite, TM was the only population that was genetically distinct from all the other populations (Table 4.2A). It is also on average more distant from the other populations (Table 4.2B, Figure 4.1B). However, the IBD plot (Figure 4.5B) does not suggest that it is more or less genetically differentiated than expected by distance alone.

Wright (1978) notes that even though F_{ST}s of less than 0.05 indicate little genetic differentiation, they are by no means negligible (Hartl and Clark 1997). The Φ_{ST} estimates of both Sequoia and Yosemite fall into this range, reflecting a relatively low degree of differentiation with a fixation index of 0.03. The lack of significant IBD within Sequoia could imply either that there is very limited gene flow between populations which are then subject to the effects of drift, or that populations are in panmixia with sufficient gene flow to homogenize the effects of local selection and drift. Given the low degree of genetic differentiation and the similarity in habitat between populations, the latter is more likely. There is ample evidence of metapopulation effects at Sequoia (Singer and Thomas 1996; Thomas et al. 1996; Boughton 1999; Boughton 2000), which suggests that the panmixia in Sequoia is likely to arise from stepping-stone roles played by multiple small populations dotted throughout the landscape, most of which were not sampled in this study.

4.3.2 GIS Modeling

None of the correlation tests between genetic differentiation and effective distance metrics for Sequoia were statistically significant (not shown). The GIS modeling for Yosemite, on the other hand, produced correlations that were all highly statistically significant ($P < 0.001$). Almost all correlations (39 out of 45 values, across all cost models and all friction values) were more than or equal to the correlation using straight line distances alone. The lowest correlation was 0.65, and the highest was 0.77. The results are summarized in Table 4.4 and Figure 4.7. The least cost path solution connecting the population TM to the other Yosemite populations is shown in Figure 4.6.

All three cost models produced effective distance metrics that were monotonically increasing with friction values. At a friction value of 11.5, all the cost models produced metrics that were better correlated with genetic differentiation than just by using straight line distances alone.

The accumulated cost metric produced Pearson correlation coefficients that ranged from 0.74 to 0.77 for all friction values and for all the cost models (Table 4.4, Figure 4.7). This compares favorably to the correlation using straight line distances of 0.66. This suggests that in all cases, this metric explained genetic variation better than straight line distances. However, the statistical significance of the differences has yet to be tested. This is due to the current lack of a readily available statistical package to assess the statistical significance of the difference of any two given Pearson correlation coefficients calculated by a permutation test.

Of the three effective distance metrics tested, the accumulated cost metric appeared to correlate better with genetic differentiation, and was also more invariant under different friction values than the other metrics under all three cost models (Figure 4.7). Under the *CostDistance* cost model, the correlation values ranged from 0.76 to 0.77. The other models produced slightly larger ranges (0.74 – 0.76 for both *PathDistance* models).

Figure 4.5B shows an IBD plot of Yosemite populations using straight line distances. Figure 4.6 shows two IBD plots of Yosemite populations using a friction value of 11.5 under the *CostDistance* model: Figure 4.6A uses the topographical distance metric, and Figure 4.6B uses the accumulated cost metric. The R^2 value for the regression line was 0.44 using straight line distance, 0.52 using the topographical distance metric, and 0.60 using the accumulated cost metric. The improved fit in the latter cases appeared to have been attributed mainly to the reduction in scatter caused by large genetic and spatial distances.

Compared to the accumulated cost metric, the path length metric produced a wider range of correlation values over different friction values. The smallest range of correlation

values was obtained using the *PathDistance(Secpower=1.0)* cost model (0.67 – 0.73). The *CostDistance* and *PathDistance(Secpower=1.0)* cost models produced a slightly larger range of values (0.67 – 0.74 and 0.66 – 0.73 respectively).

Similar to the path length metric, the topographical distance metric produced a larger range of correlation values compared to the accumulated cost metric. Figure 4.7 suggests that this metric consistently produced correlations that were lower than all the other metrics for all friction values and under all cost models.

4.4 DISCUSSION

This study is the first attempt at correlating genetic differentiation with the effects of landscape heterogeneity using information on the relative resistance posed by landscape elements. The only other published study to utilize a GIS to study genetic differentiation was done by Michels *et al.* (2001). The authors investigated whether effective geographical distance between a set of ponds interconnected by rivulets was a better predictor than simple distance of the genetic differentiation of zooplankton. A GIS was used to estimate effective geographical distances by modeling different rates of water flow between ponds. Of the three models tested, one did not improve the correlation with the observed genetic patterns, but two others produced effective distances that provided a better approximation of the rates of genetic exchange among populations.

Keyghobadi *et al.* (1999) investigated the population structure of *Parnassius smintheus* butterflies using microsatellite markers in the foothills of the Canadian Rockies. In that study, the effective distance between butterfly populations was estimated by calculating the length of probable dispersal routes along ridges between populations. This approach was not adopted for the present study because of the myriad alternative routes that are possible along ridges in topologically diverse landscapes such as Sequoia and Yosemite and because there is no reason to suppose that *E. editha* in gently-sloping terrain would migrate preferentially along ridge-tops any more than would be predicted from the nature

of the vegetation. It would be extremely difficult to identify a unique routing network which connects all the populations in such landscapes. The GIS approach used here offers a level of replicability because it does not require subjective human judgment to determine probable dispersal routes. Another advantage of this approach is the ability to conduct sensitivity analyses by testing various dispersal models.

In this study, in addition to testing three cost models, three effective distance metrics were tested. The small range of Pearson correlation coefficients produced using the accumulated cost metric – for all friction values under all the cost models – suggest that this particular metric be employed for future studies. The poor performance of the topographical distance metric was not expected: it had the most intuitive appeal among all three metrics because of its incorporation of topographical distances traversed along a least cost path between any two given populations. The path length metric, which does not incorporate topographical distances, fared slightly better than the topographical distance metric (Figure 4.7). The Pearson correlation value yielded by using topographical distance ($r = 0.65$) under all cost models fared no better than the correlation using straight line distances ($r = 0.66$) when the friction values were 6.9 and 9.2 (Table 4.4).

The accumulated cost and path length metrics performed poorly relative to the accumulated cost metric. Two possible explanations for the difference in performance are: (1) Figure 4.8 provides an example of a least cost path solution from connecting the population TM to other Yosemite populations. Such traversals along the path will almost certainly include some traversal through forested areas. The calculation of the topographical distance and path length metrics, as currently implemented, do not take into account the additional resistance afforded by forested areas. That does not, however, explain why the topographical distance metric seems to fare worse than the path length metric in all situations, even though the former incorporates information on the landscape's topology. (2) *E. editha* dispersal may not occur primarily along the least cost path. There are likely to be multiple alternate viable routes from one population to

another. A more sophisticated analysis would require the computation of alternate routes, together with some mechanism to weigh the probability of each route being actually used for dispersal.

Based on the small range of correlation values produced by the accumulated cost metric under the *CostDistance* model, it is recommended that future GIS analyses be conducted using accumulated cost metrics calculated with *CostDistance*. The friction value of 11.5 published by Ricketts (2001) appears to be suitable for studies of *Melitaeini* butterflies. The relevant resistance values for other butterfly taxa specified in Ricketts (2001), such as those for *Satyrinae*, *Argynnini*, *Pierinae*, *Polyommatini*, *Lycaenini*, can be used for similar GIS analyses of these taxa.

In Chapter 5, I investigate the association between genetic differentiation and differences in host utilization. This is done by performing a partial Mantel test between a population pairwise Φ_{ST} matrix and a host association matrix while controlling for the effects of geographical isolation. The host association matrix expresses the difference in host species utilization between populations. Geographical isolation was calculated using straight line distances. At Yosemite, no correlation between genetic differentiation and host association was found ($r = 0.32$, $P = 0.14$, Table 5.8). Yet, there was strong evidence for host-associated genetic differentiation when the same analysis was conducted at the regional scale, where populations were separated by an average of 259km. In the light of the above results, the partial Mantel test was repeated for Yosemite, but instead of using straight line distances, accumulated costs calculated using a friction value of 11.5 under a *CostDistance* model were used instead. With the use of accumulated costs to model geographical isolation, a significant association between genetic differentiation and host species association ($r = 0.42$, $P = 0.03$) was found at Yosemite.

Table 4.1: Populations included in the study.

Population Code	Population Name	Coordinates		Elevation (m)
		Latitude	Longitude	
DM	Dana Meadows	37.90756	-119.25769	3220
GDL	Gardisky Lake	37.95589	-119.25476	3227
MD	Mount Dana	37.91087	-119.23878	3482
PP	Parker Pass	37.83763	-119.20534	3403
SI	Spillway Lake	37.84435	-119.22451	3275
SL	Saddlebag Lake	37.97861	-119.28753	3111
SR	Saddlebag Ridge	37.96455	-119.25670	3445
TM	Tuolumne Meadows	37.87593	-119.36442	2620

Table 4.2: **(A) Population pairwise Φ_{ST} matrix for Sequoia (lower triangular) and Yosemite (upper triangular). (B) Straight line geographical distance matrix (kilometers) for Sequoia (lower triangular) and Yosemite (upper triangular). (C) Elevation difference matrix (meters) for Sequoia (lower triangular) and Yosemite (upper triangular).**

(A)

		Yosemite								
		DM	**GDL**	**MD**	**PP**	**SI**	**SL**	**SR**	**TM**	
Sequoia	**BB**		**0.025**	**0.024**	0.020	0.009	**0.029**	**0.014**	**0.064**	**TM**
	BR	0.018		0.001	0.037	**0.025**	0.020	0.009	**0.067**	**SR**
	CM	0.001	**0.032**		0.009	0.001	**0.021**	0.012	**0.050**	**SL**
	LBO	0.022	0.015	0.027		0.000	**0.043**	**0.035**	**0.080**	**SI**
	LOP	**0.051**	0.017	**0.048**	**0.071**		**0.037**	**0.014**	**0.063**	**PP**
	RM	**0.019**	0.013	**0.016**	**0.027**	0.022		0.004	**0.089**	**MD**
	RO	**0.067**	**0.064**	**0.066**	**0.067**	**0.096**	**0.076**		**0.057**	**GDL**
	TJ	**0.014**	**0.023**	**0.015**	0.013	**0.031**	**0.017**	**0.068**		**DM**
		TJ	**RO**	**RM**	**LOP**	**LBO**	**CM**	**BR**	**BB**	

(B)

		Yosemite								
		DM	**GDL**	**MD**	**PP**	**SI**	**SL**	**SR**	**TM**	
Sequoia	**BB**		5.37	1.70	9.02	7.60	8.31	6.33	10.02	**TM**
	BR	5.44		5.19	13.83	12.66	3.83	0.98	13.10	**SR**
	CM	12.47	14.94		8.64	7.49	8.65	6.16	11.71	**SL**
	LBO	5.15	0.77	14.18		1.85	17.24	14.79	14.63	**SI**
	LOP	15.81	11.75	16.78	11.45		15.90	13.64	12.80	**PP**
	RM	2.85	2.60	13.71	2.43	13.69		3.13	13.25	**MD**
	RO	11.64	9.39	10.91	8.77	5.95	10.33		13.65	**GDL**
	TJ	4.78	3.90	11.06	3.13	11.13	3.57	6.94		**DM**
		TJ	**RO**	**RM**	**LOP**	**LBO**	**CM**	**BR**	**BB**	

(C)

		Yosemite								
		DM	**GDL**	**MD**	**PP**	**SI**	**SL**	**SR**	**TM**	
Sequoia	**BB**		7	262	183	55	109	225	600	**TM**
	BR	225		255	176	48	116	218	607	**SR**
	CM	147	78		79	207	371	37	862	**SL**
	LBO	9	234	156		128	292	42	783	**SI**
	LOP	34	191	113	43		164	170	655	**PP**
	RM	42	267	189	33	76		334	491	**MD**
	RO	169	56	22	178	135	211		2845	**GDL**
	TJ	1	224	146	10	33	43	168		**DM**
		TJ	**RO**	**RM**	**LOP**	**LBO**	**CM**	**BR**	**BB**	

Table 4.3: Correlation between Φ_{ST} and geographical distance and elevation.

Park	Num. of Pops	Φ_{ST}	Geographical Distance		Elevation	
			Average Dist. (km)	Φ_{ST} x Dist.	Average Elev. (m)	Φ_{ST} x Elevation
Sequoia	8	**0.03** (P < 0.001)	8.7	0.24 (P = 0.25)	115	-0.02 (P = 0.51)
Yosemite	8	**0.03** (P < 0.001)	9.3	**0.66** (P < 0.001)	367	0.49 (P = 0.1)

Table 4.4: Pearson's correlation, r, from Mantel correspondence tests of population pairwise Φ_{ST} matrix against pairwise effective distance metrics. All correlations are significant at P < 0.001.

Cost Model	Friction Value	Pearson's Correlation, r		
		Accumulated Cost	Topographic Distance	Path Length
CostDistance	6.90	0.76	0.65	0.67
	9.20	0.77	0.65	0.67
	11.50	0.77	0.72	0.73
	13.80	0.77	0.72	0.73
	16.10	0.77	0.73	0.74
PathDistance (Secpower=1.0)	6.90	0.74	0.65	0.67
	9.20	0.75	0.65	0.66
	11.50	0.76	0.71	0.72
	13.80	0.76	0.72	0.73
	16.10	0.76	0.73	0.73
PathDistance (Secpower=1.5)	6.90	0.74	0.65	0.66
	9.20	0.75	0.65	0.66
	11.50	0.75	0.69	0.7
	13.80	0.76	0.72	0.73
	16.10	0.76	0.73	0.73

Figure 4.1: **(A) Populations sampled within Sequoia. (B) Populations sampled within Yosemite.**

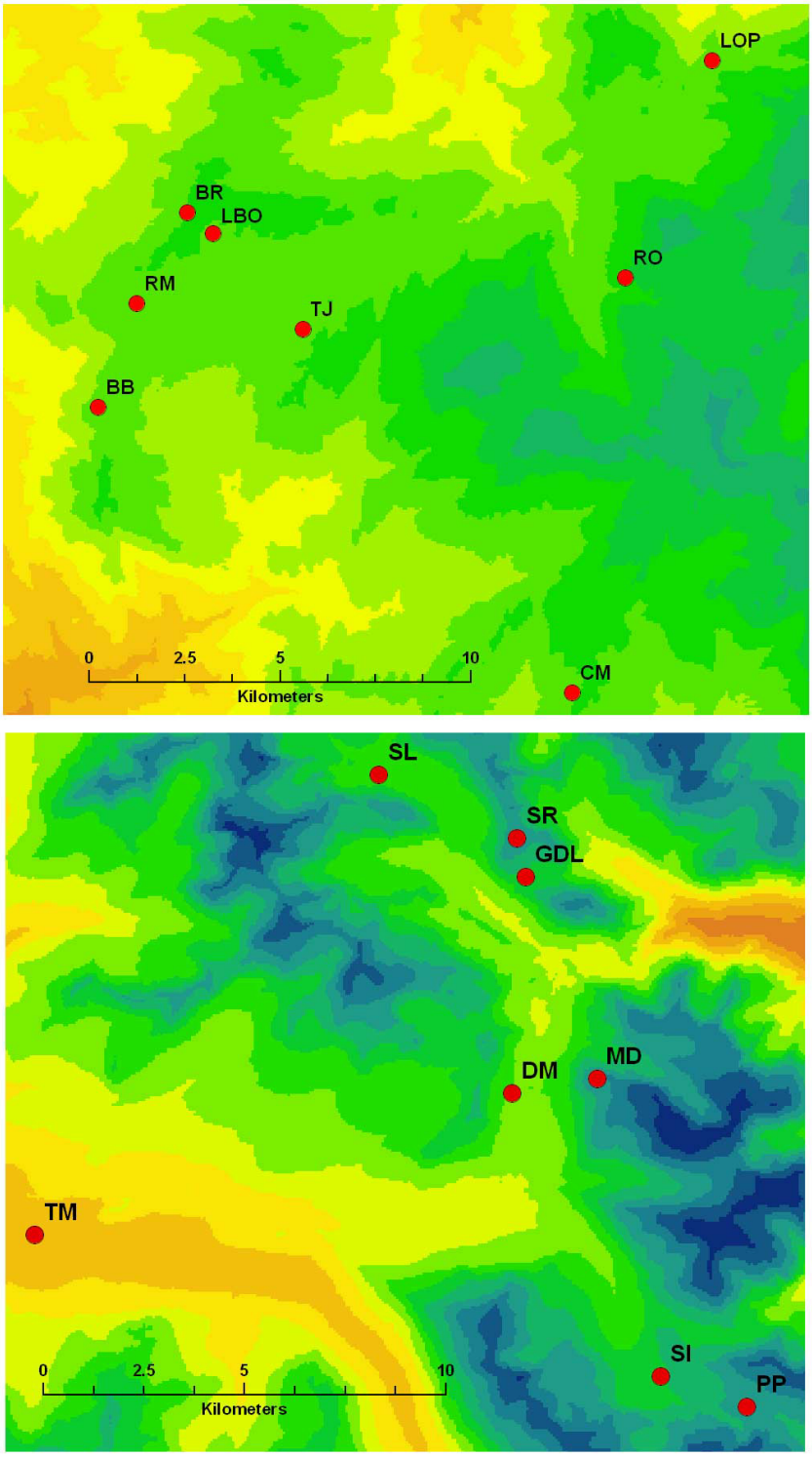


Figure 4.2: How accumulated costs are calculated.

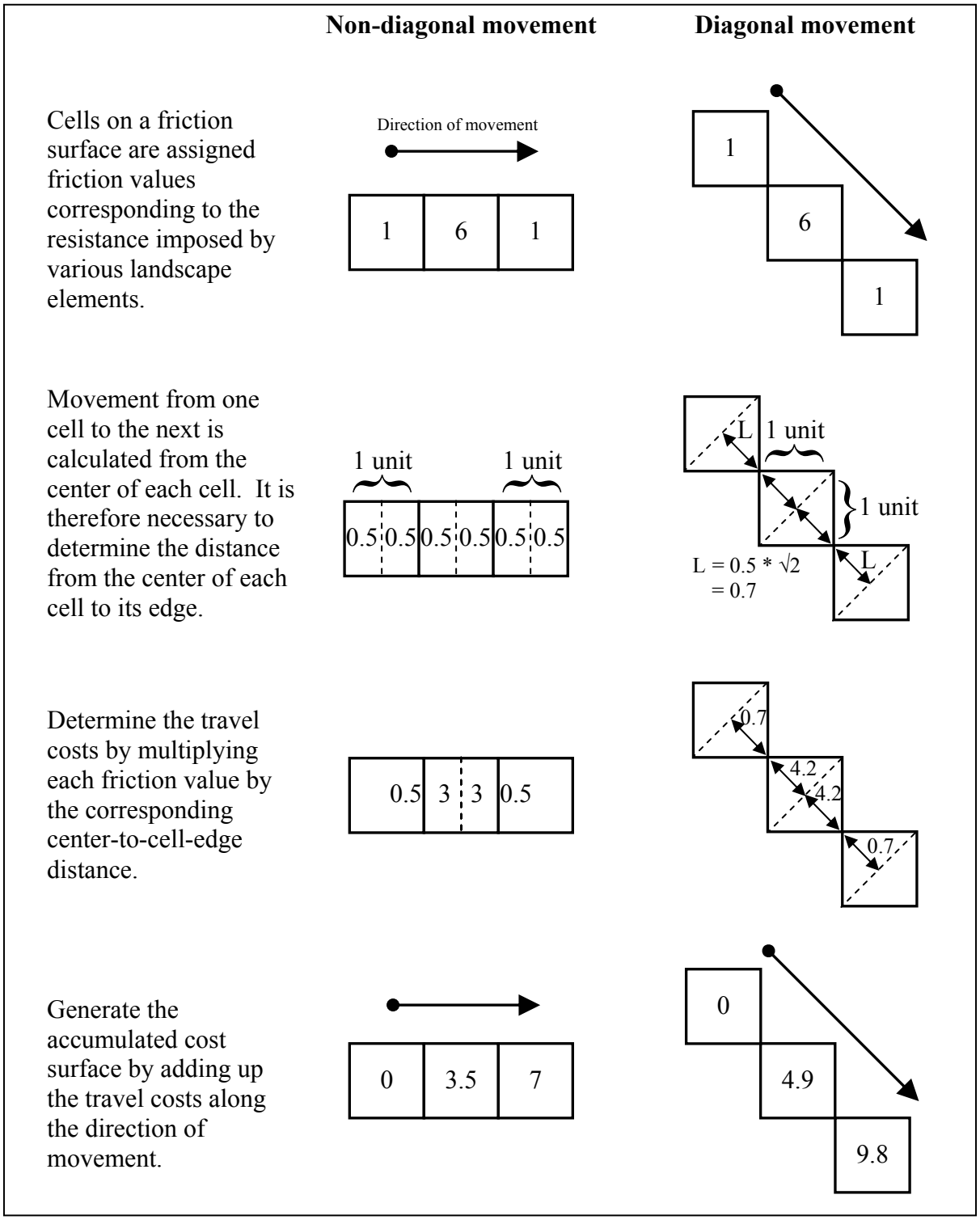


Figure 4.3 GIS methodology.

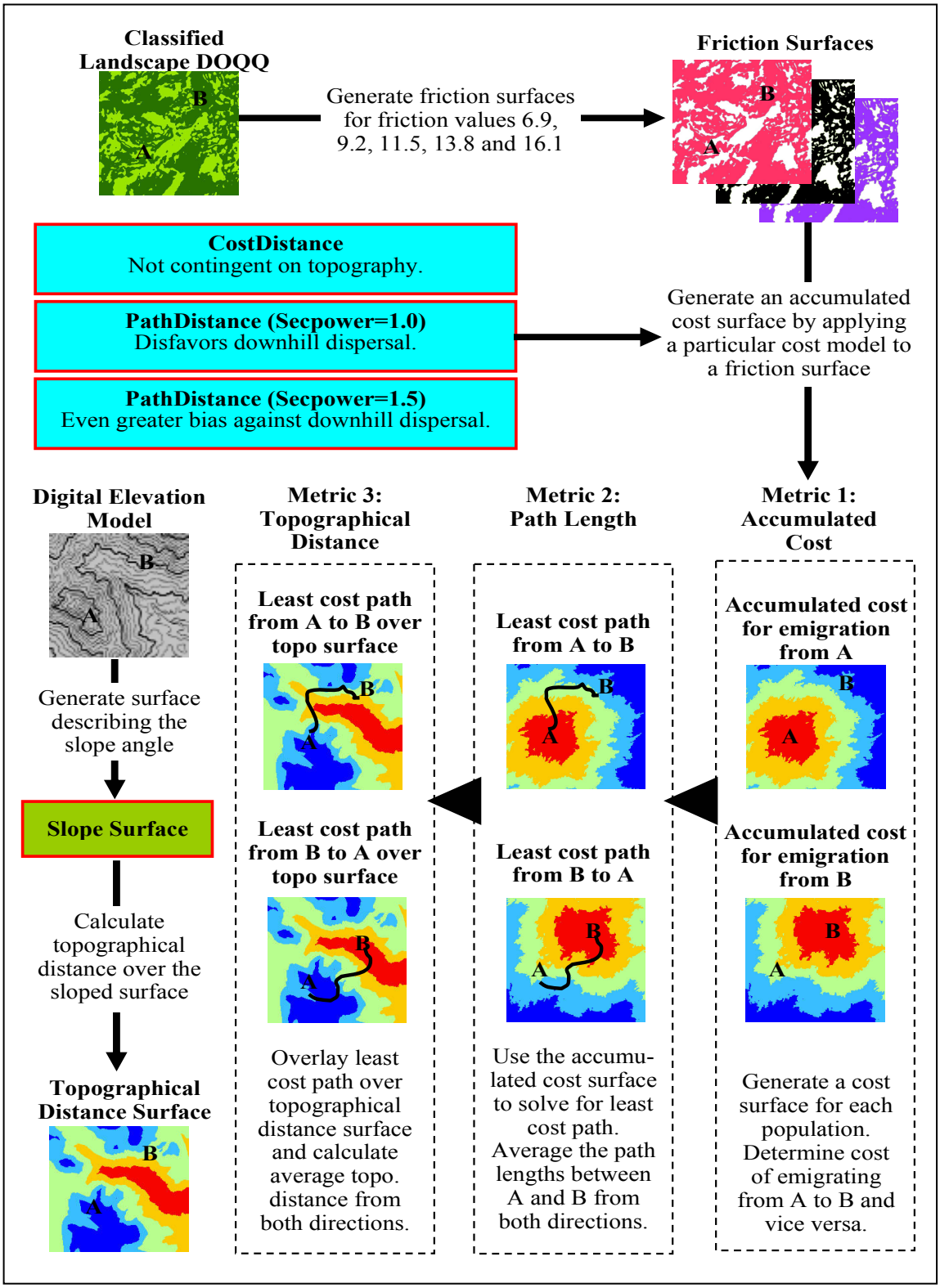


Figure 4.4: Magnitude of bias imposed on vertical movement where the bias for downhill movement is (A) secant(angle), (B) secant(angle)$^{1.5}$.

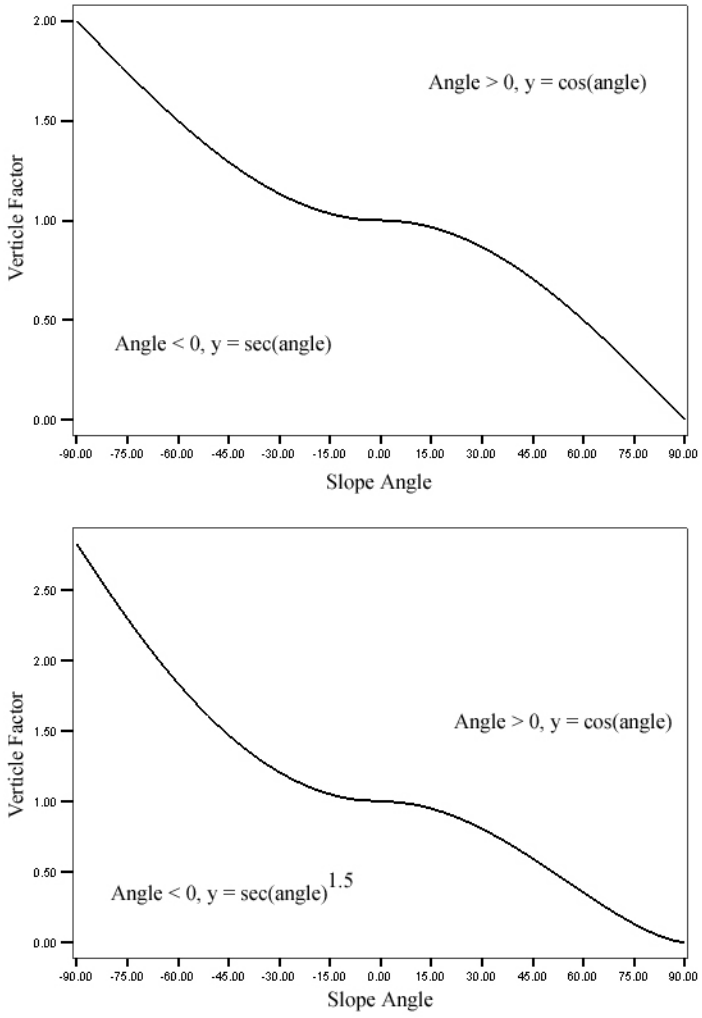
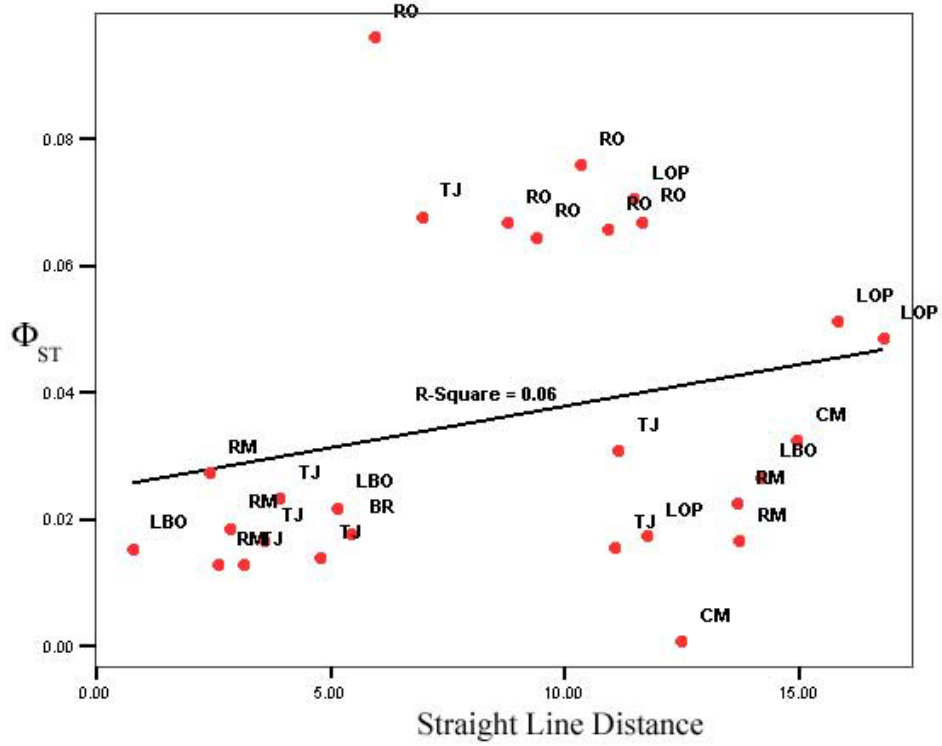


Figure 4.5: **IBD plots for (A) Sequoia, (B) Yosemite, of population pairwise distance (km) versus population pairwise Φ_{ST} estimates.**



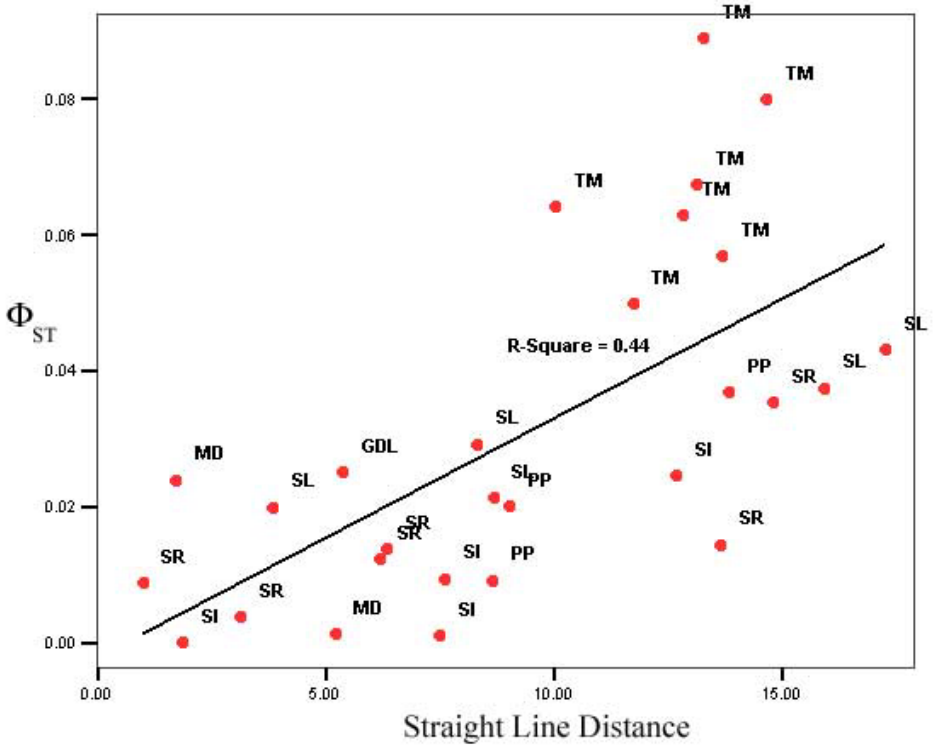
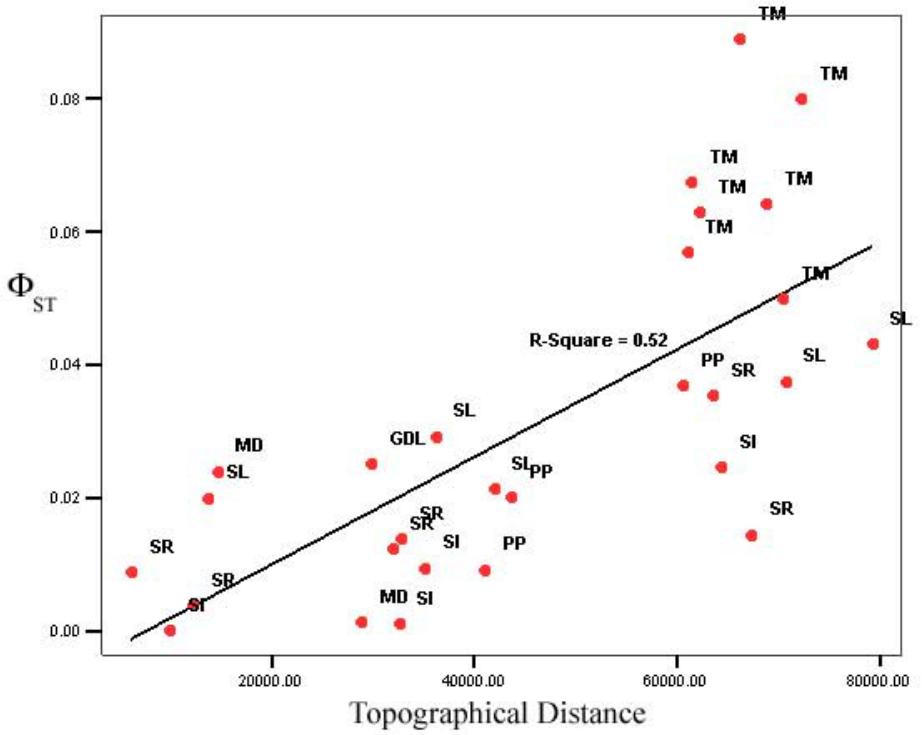


Figure 4.6 **IBD plot for Yosemite using *CostDistance*, friction value = 11.5, and (A) topographical distances, (B) accumulated costs.**



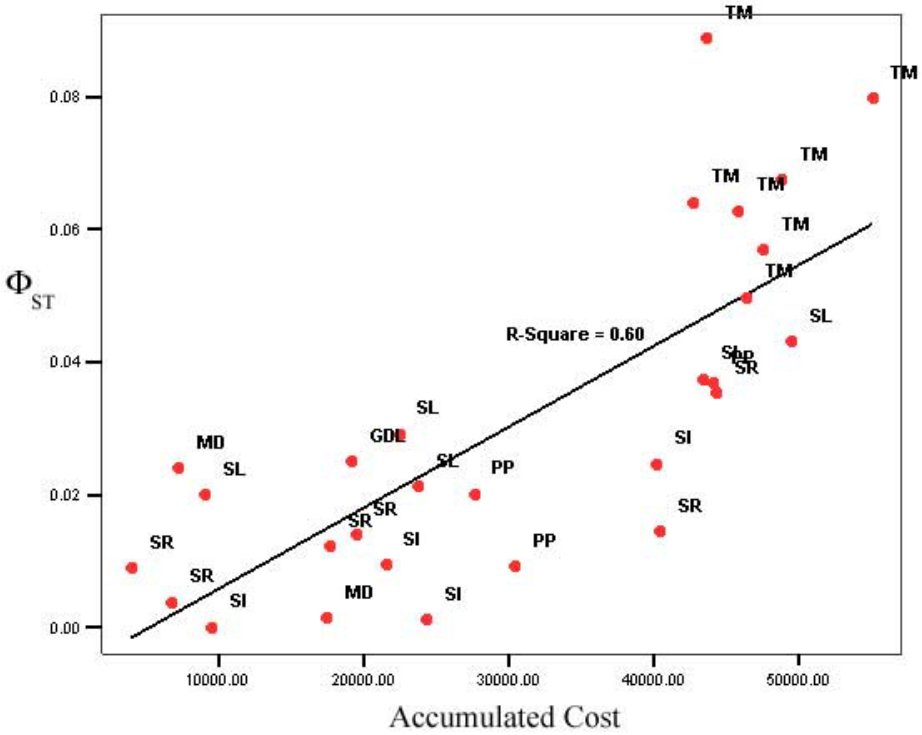


Figure 4.7: **Correlation values between population pairwise Φ_{ST} matrix and various effective cost metrics corresponding to (A) A hypothetical cost model, (B) CostDistance, (C) PathDistance(SecPower=1.0), (D) PathDistance(SecPower=1.5).**

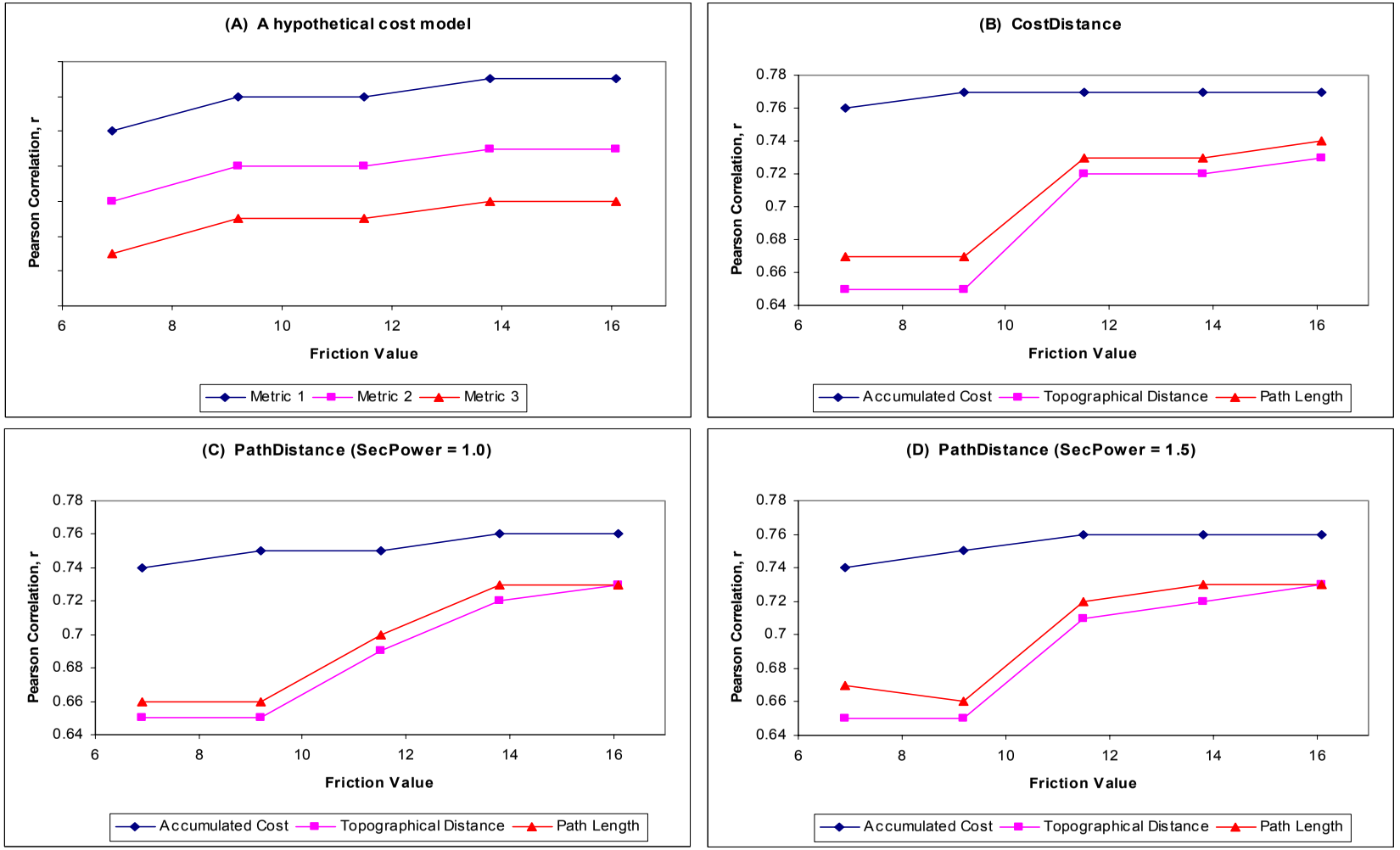
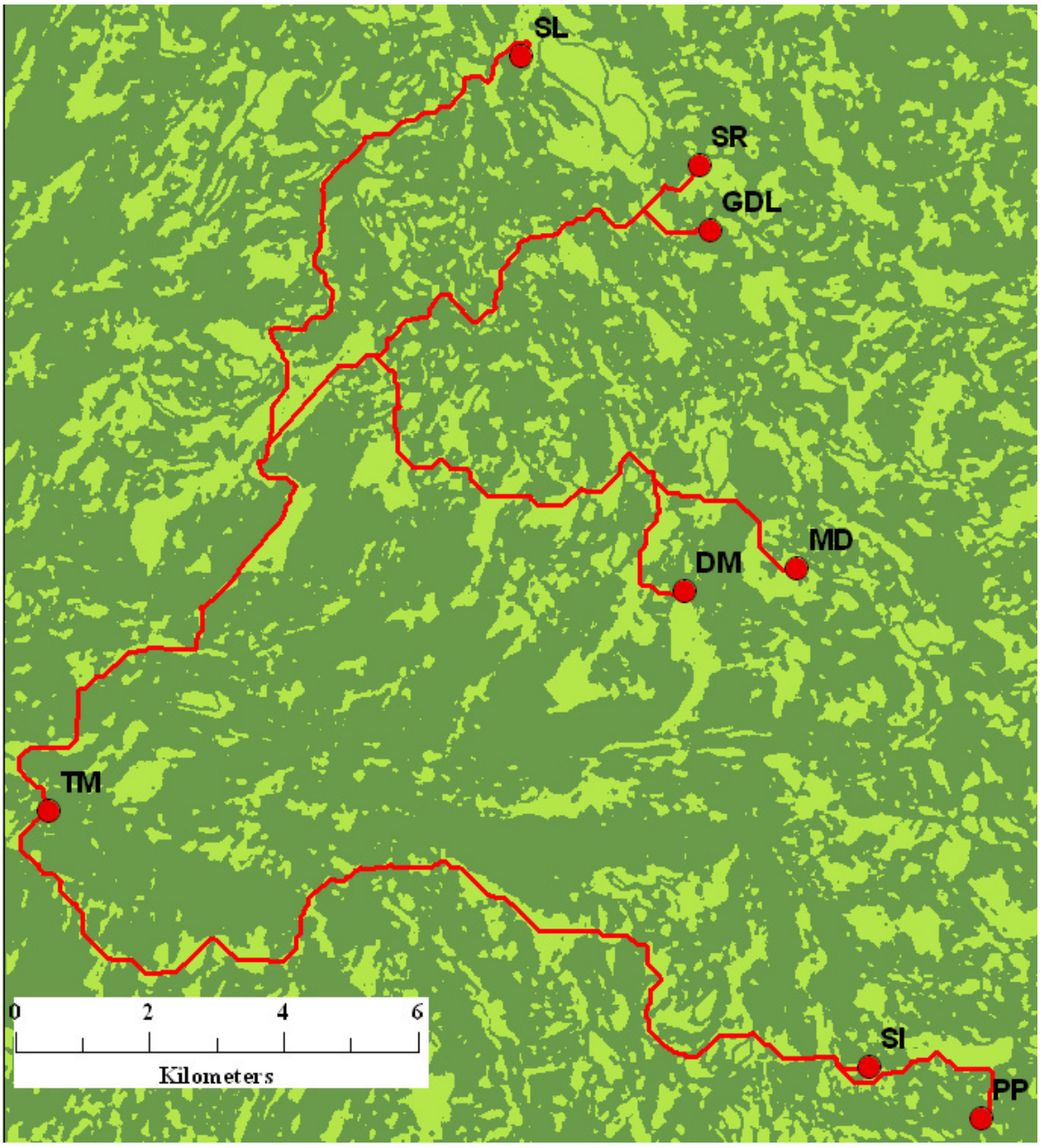


Figure 4.8: **Least cost path solution connecting the population TM to the other Yosemite populations. Light colored areas are open areas, dark colored areas are forested areas.**



References

Bonnet, E. and Y. Van de Peer (2002). "zt: a software tool for simple and partial Mantel tests." Journal of Statistical software **7**(10): 1-12.

Boughton, D. A. (1999). "Empirical evidence for complex source-sink dynamics with alternative states in a butterfly metapopulation." Ecology **80**(8): 2727-2739.

Boughton, D. A. (2000). "The dispersal system of a butterfly: A test of source-sink theory suggests the intermediate-scale hypothesis." American Naturalist **156**(2): 131-144.

Collischonn, W. and J. V. Pilar (2000). "A direction dependent least-cost path algorithm for roads and canals." International Journal of Geographic Information Systems **14**(4): 397 - 406.

Douglas, D. H. (1994). "Least-cost Path in GIS Using an Accumulated Cost Surface and Slopelines." Cartographica **31**: 37-51.

ESRI (2002). ArcGIS 8.1, ESRI Inc.

ESRI (2002). ArcGIS Spatial Analyst Functional Reference, ESRI Inc.

Excoffier, L. (2001). Analysis of Population Subdivision. Handbook of Statistical Genetics. D. J. Balding, M. Bishop and C. Cannings, John Wiley & Sonds, Ltd**:** 271-307.

Excoffier, L. (2004). Arlequin Frequently Asked Question (FAQ) List, http://lgb.unige.ch/arlequin/software/2.000/doc/faq/faqlist.htm.

Excoffier, L., P. E. Smouse, et al. (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: applicable to human mitochondrial DNA restriction data." Genetics **131**: 479-491.

Fahrig, L. (2001). "How much habitat is enough?" Biological Conservation **100**(1): 65-74.

Fahrig, L. and G. Merriam (1994). "Conservation of Fragmented Populations." Conservation Biology **8**(1): 50-59.

Hanski, I. and O. Gaggiotti (2004). Ecology, Genetics and Evolution of Metapopulations., Academic Press.

Hartl, D. L. and A. G. Clark (1997). Principles of Population Genetics. Sunderland, MA., Sinauer Associates.

Henein, K. and G. Merriam (1990). "The elements of connectivity where corridor quality is variable." Landscape Ecology **4**: 147-170.

Keyghobadi, N., J. Roland, et al. (1999). "Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus (Papilionidae)*." Molecular Ecology **8**: 1481-1495.

Levins, R. (1969). "Some demographic and genetic consequences of environmental heterogeneity for biological control." Bulletin of the Entomological Society of America **15**: 237-240.

Luetgert, J. Range: Distance/Azimuth Program, U.S. Geological Survey; http://quake.wr.usgs.gov/~luetgert/Macintosh_Software/.

MacArthur, R. and E. Wilson (1963). "An equilibrium theory of insular zoogeography." Evolution **17**: 373-387.

MacArthur, R. and E. Wilson (1967). The Theory of Island Biogeography. Princeton, New Jersey., Princeton University Press.

Michels, E., K. Cottenie, et al. (2001). "Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: a plea for using GIS modelling of the effective geographical distance." Molecular Ecology **10**: 1929-1938.

Moilanen, A. and I. Hanski (1998). "Metapopulation Dynamics: Effects of Habitat Quality and Landscape Structure." Ecology **79**(7): 2503-2515.

Norberg, U., K. Enfjall, et al. (2002). "Habitat exploration in butterflies – an outdoor cage experiment." Evolutionary Ecology **16**: 1-14.

PE-Biosystems Genotyper.

Peterson, M. (1997). "Host plant phenology and butterfly dispersal: Causes and consequences of uphill movement." Ecology **78**(1): 167-180.

Peterson, M. A. and R. F. Denno (1998). Life-History Strategies and the Genetic Structure of Phytophagous Insect Populations. Genetic Structure and local adaptation in natural insect populations. S. Mopper and S. SY, Chapman & Hall.

Ricketts, T. H. (2001). "The Matrix Matters: Effective Isolation in Fragmented Landscapes." American Naturalist **158**(1): 87-99.

Roland, J., N. Keyghobadi, et al. (2000). "Alpine Parnassius butterfly dispersal: Effects of landscape and population size." Ecology **81**(6): 1642-1653.

Schneider, S., D. Roessli, et al. (2000). Arlequin: A software for population genetics data analysis., Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Simberloff, D. S. (1974). "Equilibrium theory of island biogeography and ecology." Annual Review of Ecology and Systematics **5**: 161-182.

Singer, M. C. and C. D. Thomas (1996). "Evolutionary responses of a butterfly metapopulation to human and climate-caused environmental variation." American Naturalist **148**: S9-S39.

Thomas, C. D., M. C. Singer, et al. (1996). "Catastrophic extinction of population sources in a butterfly metapopulation." American Naturalist **148**: 957-975.

Turner, M., R. Gardner, et al. (2001). Landscape Ecology in Theory and Practice: Pattern and Process. New York, New York., Springer-Verlag.

Watt, W. B. (1969). "Adaptive significance of pigment polymorphisms in *Colias* butterflies. II. Thermoregulation and photoperiodically controlled melanin variation in *Colias eurytheme*." Proceedings of the National Academy of Sciences of the United States of America **63**: 767-774.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**: 1358 - 1370.

Weiss, S. B. and W. A.D. (1998). "Landscape-Level Phenology of a Threatened Butterfly: A GIS-Based Modeling Approach." Ecosystems **1**: 299-309.

Wright, S. (1943). "Isolation by distance." Genetics **28**: 139-156.

Wright, S. (1978). Evolution and the Genetics of Populations. Chicago, University of Chicago Press.

Chapter 5: The Role of Geographical Distance and Host Association as a Barrier to Gene Flow in *Euphydryas editha*

Abstract. Differences among allopatric populations in resource use may contribute significantly to their genetic isolation. Such effects are expected to be most pronounced in organisms showing strong local adaptation, such as the well-known adaptations of herbivorous insects to their host plants. Genetic isolation of populations is also influenced by their geographic separation. This study addresses the combined effects of local adaptation and geographic separation on population differentiation in a butterfly, *Euphydryas editha.* In addition, the effects of spatial scale on genetic differentiation are addressed by sampling at two scales: local and regional. At the local scale, where populations are separated by an average of 9km, eight populations from Sequoia National Park / National Forest (SEKI) and eight populations from Yosemite National Park (YOSE) were sampled and assayed using Amplified Fragment Length Polymorphisms. At the regional scale, where populations are separated by an average of 259km, a total of 30 populations were sampled. There was substantial genetic differentiation between populations at the regional scale ($\Phi_{ST} = 0.19$, $P < 0.001$). A smaller degree of differentiation was found at the local scale in both SEKI and YOSE ($\Phi_{ST} = 0.03$, $P < 0.001$). Geographical distance was partially responsible for genetic differentiation at the regional scale ($r = 0.55$, $P < 0.001$) and within YOSE ($r = 0.66$, $P < 0.001$), but no correlation with distance could be found for SEKI. Additionally, host association was also found to be significantly correlated with genetic differentiation ($r = 0.17$, $P < 0.01$). Populations utilizing host plant species belonging to the genus *Castilleja* (subgenus *castilleja)* were randomly scattered on a non-metric multidimensional scaling (NMDS) plot of the study populations. These populations also exhibited a strong and significant isolation-by-distance (IBD) relationship ($r = 0.7$, $P < 0.001$). This is in contrast to populations utilizing host plant species belonging to the genus *Collinsia*, which were

clumped around the origin of the NMDS plot. These populations appear to be recently colonized, and do not exhibit IBD.

5.1 INTRODUCTION

5.1.1 Background

Wright's (1943) hypothesis of genetic isolation by distance (IBD) predicts that genetic differentiation increases with geographic distance. The further a given population is from a given source of migrants, the less gene flow the population will receive from that source. In phytophagous insects, weak IBD could arise either from the effect of strong dispersal, which homogenizes allelic frequencies between populations, or it could arise from limited gene flow, which allows populations to fix for different alleles (Peterson and Denno 1998). IBD effects have been documented in various insect species (Costa and Ross 1994; Britten et al. 1995; Keyghobadi et al. 1999; Peterson et al. 2001; Wu et al. 2001). Some species show no IBD at smaller, local scales but exhibit IBD at larger scales (Peterson 1995; Peterson 1996). Peterson and Denno (1998) argue that for moderately mobile species, genetic homogeneity is achieved at small scales, while limited gene flow allows genetic differentiation over larger distances, possibly via a stepping stone model of gene flow.

Peterson and Denno (1998) also note that the majority of studies report a lot of scatter in the IBD relationship studied. Such scatter can either be attributed to populations or individuals which are either more differentiated than expected by distance alone, or less differentiated than expected by distance alone. They attributed this scatter to processes like genetic drift, extinction – recolonization processes, or a combination of both, which weakens the effects of IBD. Potential causes for such "noisy" IBD relationships also include local adaptation of resource utilization by conspecific populations.

In addition to IBD, the use of different diets by allopatric populations may contribute significantly to their genetic isolation. Local adaptation, in addition to spatial isolation and physical landscape barriers, may disrupt the association between gene flow and distance by acting as a barrier to gene flow through behavioral (Kawecki 1998; Berlocher 1999) or physiological traits. Any such effects are expected to be most pronounced in organisms showing strong local adaptation to diet, such as herbivorous insects. Herbivorous insects are plant specialists, and are therefore prone to strong disruptive selection pressures (Mopper and Strauss 1998). Host-associated genetic differentiation has been attributed to host fidelity, defined by Feder (1998) as the tendency of individuals to seek mates on the host species on which they themselves developed (Feder et al. 1993; Feder et al. 1994; Via 1999; Abrahamson et al. 2001). In addition to host fidelity, Funk et al. (2002) stress the role of two other premating barriers in ecologically driven speciation: temporal isolation associated with differences in timing of host plant development, and behavioral isolation associated with courtship signals or mating preferences. Genetic differentiation has been associated with temporal isolation (Wood and Keese 1990; Feder et al. 1993; Feder and Filchak 1999; Thomas et al. 2003) as well as behavioral isolation (Abrahamson et al. 2001; Thomas et al. 2003).

In contrast to species known to show host-associated genetic differentiation in sympatry, the subject of the present study, the butterfly *Euphydryas editha,* exhibits neither host fidelity nor substantial behavioral isolation associated with mating. There is evidence for temporal effects of host senescence (see below, "Effect of larval performance on isolation"), but this has so far only been observed at one site. The absence of such barriers argues against finding host-associated genetic differentiation in *E. editha*. On the other hand, *E. editha* are locally adapted to the plants in their habitats. Known adaptations include adaptive oviposition choice by adults (Thomas and Singer 1998) and larval performance traits that facilitate survival on host plants selected by adults (Thomas and Singer 1998, Rausher 1982). Such adaptations may isolate monophagous populations on different hosts. In addition, because populations have varying diet

breadths, specialists with narrow diet-breadths should be less able than generalists to find suitable habitats that they can use (Futuyma and Moreno 1988). The combined effects of maladaptive oviposition behavior and lower larval performance on a foreign host may prevent immigrants from contributing genes to recipient populations, allowing differentiation over time through drift, especially because population bottlenecks are frequent (Ehrlich 1965; Singer et al. 1993; Thomas et al. 1996). Therefore, if host-associated genetic differentiation were to be found in *E. editha,* it would have developed in the presence of substantial local adaptation but in the absence of the host fidelity that some authors (Feder 1998; Mopper and Strauss 1998) would regard as its primary cause.

Objectives. The first objective of this study is to elucidate the mode of gene flow in the checkerspot butterfly *Euphydryas editha*: is there moderate gene flow at all spatial scales, or is there high gene flow at smaller scales with little gene flow at larger scales?

The second objective of this study is to ask if we can find host-associated genetic differentiation in allopatric populations of the butterfly *Euphydryas editha*. If so, to what extent is the degree of isolation between populations associated with differences in host use after taking into account the effects of spatial isolation?

5.1.2 Study system

E. editha occurs in diverse habitat types, including chaparral, oak woodland, serpentine grassland, montane coniferous forest, and high elevation tundra (Figure 5.1). *E. editha* uses a diversity of host plants in the families *Scrophulariaceae, Plantaginaceae, Antirrhinaceae,* and *Orobanchaceae*. Populations in the Sierra Nevada utilize five different host genera for oviposition. 70% of these populations are monophagous, while 30% use from 2 – 4 host genera. Monophagous populations on one host genus often occur interdigitated with populations on a different host. This spatial variation of insect diet is driven mostly by genetic variation for oviposition preference (Singer and Parmesan 1993). Population structure is also variable within this species: those found in

Sequoia National Forest and Yosemite National Park typically exist as large metapopulations, while in some regions most of the insects exist in well-defined and relatively isolated populations.

The genus *Castilleja* is the presumed ancestral host plant for *E. editha*, utilized as the principal host by populations east of the Sierra Nevada. In the Great Basin and the Rocky Mountains, some of the known hosts include *Castilleja linariiafolia*, *C. chromosa*, and *C. lapidicola* (Britten et al. 1995). *Castilleja* is also the principal host of *E. chalcedona*, the sister species to *E. editha* (Brussard et al. 1985) (but see Wahlberg and Zimmermann 2000; Zimmermann et al. 2000).

The other host genera that are used as principal hosts of *E. editha* populations included in this study are *Collinsia*, *Pedicularis*, *Penstemon*, and *Plantago*.

Effect of Adult Preference on Isolation. Host preference of checkerspot butterflies is known to influence their patterns of dispersal. Thomas and Singer (1987) showed that adaptive oviposition preference reduces movements of female *E. editha* between populations adapted to different hosts. In the related species *Melitaea cinxia* (Glanville fritillary), this process biases colonization of empty habitat patches towards those patches whose host composition matches the local adaptation of migrating females (Hanski and Singer 2001). The reverse effect also occurs: movement of checkerspots influences differentiation among populations in host preference, with the result that the relative use of two hosts by *M. cinxia* populations could not be predicted from the relative abundance of the hosts in the habitat alone (Kuussaari *et al.* 2000). Over spatial scales of 1 – 4km, *E. editha* populations using different hosts are not so strongly isolated as to enable them to evolve independently in terms of host adaptation (Singer and Thomas 1996) even though effects of gene flow on preference were detected at this scale. However, populations 20 km apart may have no overlap of host preference. Variation among *E. editha* populations in host-related adaptations is stronger than variation within them (Rausher 1982; Singer and Parmesan 1993; Singer 1994). This is expected to restrict the successful colonization

of new habitat patches by migrant females. There is also evidence of behavioral adaptation of adults to the host plants. Adults in a rapidly evolving population (Rabbit Meadow, Sequoia National Forest) exhibited a suite of maladaptive oviposition behaviors when attempting to oviposit on a novel host (Singer 1994).

Effect of Larval Performance on Isolation. *E. editha* larvae are often physiologically adapted to feed on their locally traditional host (Thomas and Singer 1998), and perform poorly on the hosts of other conspecific populations (Rausher 1982). In a metapopulation of *E. editha* using two host genera, *Collinsia* and *Pedicularis* in separate patches, larvae on *Collinsia* developed faster, generating adults ten days earlier than larvae on *Pedicularis* (Singer 1983; Boughton 1999). The result of this host-associated difference was that adults developing on *Pedicularis* normally emerged so late that they were unable to produce surviving offspring in patches of *Collinsia* because of host senescence. Insects developing on *Collinsia* had high offspring survival on both hosts.

Effect of Preference-Performance Correlation. Correlations between maternal preference and offspring performance appear to be absent from most insect populations (Jaenike 1990; Fox 1993) (but see Via 1986; Hawthorne and Via 2001). However, within populations of *E. editha,* mothers with particular oviposition preferences produced offspring with particular performances (Ng 1988; Singer et al. 1988). In one population where a novel host had recently been incorporated into the diet, insects that strongly preferred the traditional host produced larvae that grew more slowly on the novel host (Ng 1988). Similar associations occur among populations: where it has been investigated, interpopulation variation of performance parallels that of preference (Rausher 1982) (also M.Singer & C.D. Thomas, unpublished data). This implies that a female immigrating to a patch that does not contain her preferred host species may incur low offspring success if she oviposits on the host used by the local butterflies.

5.2 MATERIALS AND METHODS

5.2.1 Sampling

In order to compare the effects of host association at different geographical scales, the collection of specimens was conducted at the local level and at the regional level. At the local level, where populations are separated by an average of 9km, specimens were collected from eight populations within Sequoia National Park / National Forest (denoted SEKI), and eight populations within Yosemite National Park (denoted YOSE).

At the regional scale, where populations are separated by an average of 259km, 26 populations were sampled. Two populations each from SEKI and YOSE were included in the regional analysis, for a total of 30 populations. Populations within SEKI utilize either *Pedicularis semibarbata* or *Castilleja applegatei* (subgenus *Castilleja*). One population, TJ, utilizes both hosts. The others utilize *Pedicularis* or *Castilleja* almost exclusively. Rabbit Meadows (RM) and Rowell Meadows (RO), which utilize *Pedicularis semibarbata* and *Castilleja applegatei* (subgenus *Castilleja*) respectively, were chosen to represent SEKI.

In the case of YOSE, populations utilize Castilleja nana (subgenus Colacus), Castilleja lemmonii (subgenus Colacus), or Castilleja miniata (subgenus Castilleja). A distinction was made between populations which utilize *Castilleja* subgenus *castilleja* and *Castilleja* subgenus *colacus*, henceforth referred to as *Castilleja* and *Colacus* respectively. The reason for making this distinction is that the butterflies themselves clearly do so. At sites where species of both sub-genera occur in sympatry, *E. editha* populations have used one but not both. Accordingly, the populations Tuolumne Meadows (TM) and Saddlebag Lake (SL) were selected to represent YOSE. TM and SL utilize as their respective primary hosts *Castilleja lemmonii* (subgenus *Colacus*) and *Castilleja miniata* (subgenus *Castilleja*) respectively.

The total number of specimens assayed for this study is 498 specimens from 42 populations, for an average of 12 samples per population. The populations examined in this study are listed in Table 5.1, and their geographical relationships are depicted in Figure 5.1.

Both adult insects and larvae were used for the molecular assay. Larvae were collected from the field between 1999 and 2003. In populations where the adults were known to oviposit on more than one host species, an attempt was made to collect an equal number of larval specimens from each host species. These specimens were labeled with the host plants they were collected from and frozen at -80 °C. Freshly dead or dying adults were similarly preserved. In order to minimize the impact of sampling on certain populations, adults were caught and released after a small wing clipping was taken. The clippings were preserved in 100% EtOH for the molecular assay.

5.2.2 AFLP Assay

Sample preparation. DNA was isolated from specimens either frozen at -80 °C or preserved in 100% EtOH. For adult specimens, a single leg was removed for DNA isolation, and the rest of the body stored for future use. In order to minimize possible contamination from parasitoids in the larval body, DNA was extracted from the heads of larvae, which are less likely to contain parasitoids. Wing clippings were completely consumed in the extraction process.

DNA extraction. Qiagen DNeasy Tissue kits (Catalog No. 69506) were used for DNA extraction. The Qiagen protocol was followed except for the following four modifications: (A) A CTAB extraction buffer (10% 1 M Tris-HCL pH 8.0, 28% 5 M NaCl, 4% 0.5 M EDTA pH 8.0, 2% CTAB) was used instead of Qiagen's tissue buffer. (B) Prior to extraction, 2 µl of beta-mercaptoethanol per 1ml of extraction buffer was added to the CTAB buffer. (C) Samples were crushed in a 1.5ml Eppendorf tube in

liquid nitrogen, and then homogenized in 200 µl of the extraction buffer. (D) The DNA was eluted in two 100 µl washes of buffer AE for a final DNA volume of 200 µl.

AFLP PCR reaction. The Applied Biosystems (ABI) AFLP protocol for small plant genomes (www.appliedbiosystems.com; protocol 4303146) was used with the following six modifications: (A) 30 ng of DNA was used for restriction-ligation and incubated using a thermocycler with a heated lid for 3 hours. (B) 120 µl of $TE_{0.1}$ was added to the restriction-ligation product instead of the recommended 189 µl. (C) Pre-selective amplification with the regular genome preselective primer mix was next performed in 10 µl reactions. (D) Instead of the recommended 20 cycles of denature – anneal – extension, 25 cycles were used instead. (E) The pre-selective product was not diluted for the selective amplification. (F) For the selective PCR, the 25 cycles of denature – anneal – extension (20 s at 94 °C, 30 s at 56 °C, 2 min at 72 °C) were used instead of the recommended 20.

The three primer pairs from ABI used for this study were: (1) EcoRI –ACA / MseI – CAT, (2) EcoRI – ACA / MseI – CTG, and (3) EcoRI – AAG / MseI – CTG. 3 µl of the selective product was mixed with 0.4 µl of GeneScan 500 Rox and 6.6 µl of Hi-Di Formamide. Samples were processed on an automated ABI 3100 Genetic Analyzer.

5.2.3 Analysis

Scoring of AFLP data. Raw fragment data from the 3100 Genetic Analyzer were scored in Genotyper v3.6 (PE-Biosystems). A custom written Genotyper macro was used to export the scored data to a custom written PC program, AFLPal (see Chapter 2 "An efficient AFLP protocol for butterflies: from laboratory to analysis"). In most cases, a given locus for an individual is scored as either absent or present, depending on whether an AFLP fragment was generated for that locus. However, certain populations were characterized not only by the absence or presence of a fragment, but also by the number of copies of that fragment. This was reflected by the fragment peak height in the

Genotyper software. Peak heights with bimodal frequency distribution reflect individuals from two different populations and should be scored as such (personal observation). AFLPal was used to analyze the frequency distribution of the number of fragments associated with any one locus. Based on that analysis, the program allows the user to control how the absence-presence matrix is to be generated. In addition, AFLPal allows the user to specify parameters to denote a fragment that is present (scored as '1'), absent (scored as '0'), or ambiguous (scored as '?'). The latter state can be used by many programs to denote missing data.

Relationship between populations. Arlequin (Schneider et al. 2000) was used to analyze the absence – presence matrix generated by AFLPal. The matrix of Nei's corrected average pairwise differences (Nei and Li 1979) calculated by Arlequin was used to generate a non-metric multidimensional scaling (NMDS) plot using NCSS (Hintze 2001) to visualize the relationships between populations. NMDS is an ordination technique which detects nonhierarchical geographical structure by reducing the multi-dimensional relationship between entities to a smaller number of dimensions, typically two or three. UPGMA, on the other hand, is a clustering technique which has been widely applied in the past to both inter-specific and intra-specific studies. UPGMA, unlike NMDS, has the benefit of providing information on hierarchical evolutionary pathways (Guiller et al. 1998), but has the limitation of not being able to uncover non-recticulate and non-hierarchical patterns (Lessa 1990). Despite this major limitation, the AFLP data was analyzed using UPGMA in PAUP* (Swofford 1998) to provide a common basis for comparison with previous population genetic studies done on *E. editha*. The same genetic distance matrix used for the NMDS was used as input to the UPGMA clustering algorithm.

Population structure. The AMOVA (Excoffier et al. 1992) module in Arlequin was used to determine Φ statistics, equivalent to the Weir and Cockham's (1984) θ statistics (Excoffier 2001). This allowed for the quantification of amount of genetic variation

within and between populations, as well as between groups of populations on different hosts. All Φ statistics were computed for statistical significance with 5040 permutations.

Geography associated structure. The population pairwise Φ_{ST} matrix calculated by Arlequin was used in full and partial Mantel correspondence analyses to calculate r, the Pearson correlation coefficient, using the program ZT (Bonnet and Van de Peer 2002). All significance tests with ZT were performed with 10000 permutations. To test for effects of IBD, the correspondence test was done between the Φ_{ST} matrix and a pairwise geographical distance matrix between populations. This test was done at both the local and regional geographical scales. The program Range (Luetgert, USGS) was used to calculate the geographical separation between populations measured as straight line distances between two points. Geographical coordinates were acquired from on-site readings using a hand-held GPS receiver or estimated from a topographical map. The elevation above sea level for each population was determined in a similar fashion. Differences in elevation may act as a barrier to gene flow, mediated either through behavior (such as a tendency to fly uphill versus downhill), physiology (such as operational constraints of flight muscles related to ambient temperature), or host phenology (such as host senescence). To test whether such effects are present, Mantel correspondence tests were conducted between the Φ_{ST} matrix and a matrix representing differences in elevation between populations.

Host-associated structure. To test for effects of host association, partial Mantel tests were conducted between the Φ_{ST} matrix and host association matrix while holding the effects of geography constant. This correspondence test was done at the host species, genus, and sub-genus levels, as detailed in section 5.2.4, "Host Utilization Matrix", below.

To examine if host utilization was associated with degree of gene flow, populations were grouped by their principal hosts at the subgenus level and pairwise Φ_{ST} matrices computed to reflect the relationships between populations on a given host. Mantel tests were conducted between these matrices and the corresponding geographical distance

matrix. To examine if there was any detectable subdivision between populations on any two given hosts, AMOVA was used to analyze the amount of genetic variation between groups of populations on those hosts.

Genetic diversity estimates. AFLPSurv (Vekemans 2002) was used to calculate Lynch & Milligan's (1994) *Hw* gene diversity estimate. *Hw* is the average heterozygosity estimate for a given set of populations, where heterozygosity is calculated using equation 4a after Lynch and Milligan (1994) to take into account the frequency of the null allele. Methods which assume that the frequency of the dominant allele is equal to the AFLP fragment frequency may overestimate the frequency of the marker allele. This is because AFLP markers are dominant markers, and it is not possible to distinguish between dominant homozygotes and heterozygotes when determining the AFLP fragment frequencies of diploid individuals. Fragment frequencies determined in this manner therefore provide inaccurate estimates of gene diversity. AFLPSurv employs using Zhivotovsky's (1999) Bayesian method to estimate the null allele frequency, which gives a more unbiased estimate than just using the AFLP fragment frequency. Among the statistics reported by AFLPSurv are the frequency of the marker allele, the frequency of the null allele, and the variance of the frequency of the null allele. These figures were used to calculate the average heterozygosity for a given population in Excel using equations 4a and 5 after Lynch and Milligan (1994). *Hw* for a given group of populations on a given host is the average of the heterozygosity estimates of each population (denoted *H*) in that group (equation 7, Lynch and Milligan 1994). To test for statistical significance between average *Hw* estimates for two groups of populations on any two given hosts, Minitab (Minitab Inc.) was used to run a Mann-Whitney on the two sets of population level heterozygosity (*H*) estimates.

5.2.4 Host Utilization Matrix

Most *E. editha* populations are monophagous on a single host species. Polyphagous populations use from 2 – 4 host genera. In order to analyze the correlation between

genetic differentiation and differences in host use, a host utilization matrix was created for use in Mantel correspondence tests. This matrix summarizes the differences in host utilization between populations.

To create this matrix, an Excel spreadsheet with populations listed as rows and host plants listed as columns was created (Table 5.1). For a given population, '1's were marked in the relevant cells to represent utilization of a particular host for oviposition (Singer, M.C. personal communications). Populations that are polyphagous may show variation in host use between years (Singer et al. 1993; Singer and Thomas 1996; Singer 2004). For any given population where field collected larval specimens were used for the AFLP assay, a '1' was marked for each host species from which larvae were collected. If adults were used for the assay but no larvae were collected (as is the case when old frozen specimens are used, or when wing clippings of live adults were used in order to minimize possible demographic impacts), a population was deemed to utilize a certain host if at least 5% of egg clusters found in the field were attributed to that host for that year (or in recent years, if no information were available for that particular year).

Three host utilization matrices were created for analyses: one at the species level (Table 5.6), another at the genus level (not shown), and another at the sub-genus level (Table 5.6) with *Castilleja* and *Colacus* as distinct entities.

PopTools (Hood 2003), an Excel plug-in utility for the analysis of matrix population models, was then used to generate a host-utilization distance matrix using the Sorenson coefficient. Each cell in this matrix represents the degree of difference in host utilization for oviposition between populations. Sorenson's coefficient is often used to compare species diversity between sites in community ecology. In this study, the coefficient ranges from a value of zero which suggests no difference in host utilization between sites, to a value of one which suggests non-overlapping host utilization between sites. The coefficient weighs more heavily the presence of a species than its absence. Its use in this

study reflects the possibility that the lack of use of a host at any one given site may be a sampling artifact (Krebs 1989).

5.3 RESULTS

5.3.1 Relationship between Populations

A total of 547 AFLP loci were identified, 546 which were polymorphic. The primer pairs EcoRI –ACA Fam / MseI – CAT, EcoRI – ACA Fam / MseI – CTG, and EcoRI – AAG Joe / MseI – CTG generated 255, 132, and 160 markers respectively. Table 5.2 shows the pairwise Φ_{ST} matrix generated by Arlequin for the populations at the regional scale. Of the 435 Φ_{ST} values, five were negative, ranging from -0.066 to -0.002. These were attributed to pairwise comparisons between BLC and five other populations: YP, CG, TR, DP, and IOH. Negative values can sometimes occur in the absence of genetic structure when the true value of the fixation index is zero (Excoffier 2004). The negative Φ_{ST} values were converted to zero for further analysis. Table 5.3 shows the corresponding matrices for SEKI and YOSE.

Figure 5.2 shows the results of the NMDS analysis. Populations are color coded by the principal hosts as detailed in section 5.2.4, "Host Utilization Matrix". Table 5.4 shows the first ten eigenvalues and cumulative percentage explained by each ordination axis. Axis #1 accounts for 26% of the total variation, while axis #2 accounts for a further 13%, for a total of 39%. Three dimensions are required to account for 50% of the total genetic variation, while up to six dimensions are required to account for 69% of the total variation.

Geography-associated structure visible in the NMDS plot. Formal analyses of geographic structure are given in Section 5.3.2 "Geography associated population structure". A preliminary sense that geography is important in population differentiation can be gleaned from visual examination of Figure 5.2. Despite the relatively low

percentage of variation accounted for by axis #1, it nevertheless captures a component of the geographical association between populations. With the exception of the populations DP, MUD and IOH, the populations on the positive side of axis #1 are all found on the eastern flank of the Sierra Nevada. DP is situated to the west of the Central Valley, but falls out near the zero point on axis #1. MUD and IOH, which also fall out near the zero point, are populations on the western Sierra foothills. The populations WH and SK, while situated at the southern-most edges of this study, fall out on the negative side of axis #1, together with the western Sierra populations.

Host-associated structure visible in the NMDS plot. Formal analysis is given in Section 5.3.3 "Host-associated population structure". Examination of the figure reveals a component of the genetic signal that is host-associated. Populations utilizing *Collinsia* are clustered around the origin, while populations utilizing the genus *Castilleja*, suggested from mtDNA analysis as the ancestral host plant genus of *E. editha* (Radtkey and Singer 1995), appear to be more scattered. In particular, populations that utilize the subgenus *Castilleja* appear to be more scattered than populations that utilize subgenus *Colacus*. This host signal is also apparent in the UPGMA solution, depicted in Figure 5.3. This solution exhibits concordance with the NMDS solution in several aspects. In both solutions, *Collinsia* feeders form a distinct cluster. However, the geographic relationship separating eastern and western Sierra populations apparent in the NMDS solution is not evident in the UPGMA solution.

The same pattern of scatter among genotypes on subgenus *Castilleja* was also found within a particular population in the SEKI system. This population, TJ, utilizes *Pedicularis semibarbata* and *Castilleja applegatei*. Figure 5.5 shows the NMDS plot for larvae collected on these two hosts at TJ in the year 2000. Further analysis is given in Section 5.4.4 "Differentiation at the local scale and its correlates".

5.3.2 Geography-Associated Population Structure

Regional scale. Overall Φ_{ST} at the regional scale for the 30 populations was estimated at 0.19 ($P < 0.001$), with 81% of the total genetic variation found within populations. The mean elevation difference between populations is 1013m. For the set of populations used in the regional analysis (including only 2 populations each for SEKI and YOSE) the minimum, maximum, and mean distances between populations, measured as the crow flies, are 10.3km, 858.7km, and 259km respectively. If the Great Basin population PEQ were excluded, the minimum, maximum, and mean distances between populations would be 10.3km, 814.9km, and 234.7km respectively.

Mantel correspondence tests between pairwise Φ_{ST} and geographical distance matrices detected a strong IBD effect ($r = 0.55$, $P < 0.001$) (Figure 5.4). After controlling for the effects of host association at the species and sub-genus level (see Section 5.3.3 "Host-associated population structure"), the effects of spatial isolation were still significant ($r = 0.56$, $P < 0.001$ and $r = 0.55$, $P < 0.001$ respectively). There was no evidence that elevation differences between populations accounted for any of the genetic variation at this scale ($r = -0.04$, $P = 0.27$). These results are summarized in Table 5.7.

Local scale. The eight populations within SEKI demonstrated a low degree of genetic differentiation. Φ_{ST} was estimated at 0.03 ($P < 0.001$), with 97% of the variation found within populations. The minimum, maximum, and mean distances between populations are 0.8km, 15.8km, and 8.7km respectively. The mean difference in elevation between populations is 115m. No significant relationship was found between genetic differentiation and geographical distance or elevation (Table 5.7).

Similarly, the eight populations within YOSE demonstrated a low degree of genetic differentiation. Φ_{ST} was estimated at 0.03 ($P < 0.001$), with 97% of the variation found within populations. The minimum, maximum, and mean distances between populations are 1.7km, 17.2km, and 9.3km respectively. The mean elevation difference is 367m.

There is no evidence that genetic structuring is correlated with elevation. However, geographical separation is correlated with genetic structuring ($r = 0.66$, $P < 0.001$). These results are summarized in Table 5.7.

5.3.3 Host-Associated Population Structure

Regional scale. At the regional scale, genetic structure was correlated with differences in host species, after controlling for the effects of geography ($r = 0.24$, $P < 0.001$). There was also evidence for host effects at the sub-genus level, although the association was weaker ($r = 0.17$, $P < 0.01$). This relationship disappeared when analyzed at the genus level ($r = 0.05$, $P = 0.16$), where the subgenera *Castilleja* and *Colacus* were analyzed as the genus *Castilleja*. The results are summarized in Table 5.8.

Populations were grouped according to the principal host utilized by each population. Each population group was then analyzed for evidence of IBD. The same population groups were used for the AMOVA and genetic diversity analyses. Only four populations utilized *Plantago*, and three utilized *Penstemon*. Due to the small sample sizes, these hosts were omitted from the analysis, leaving four hosts: *Castilleja, Colacus, Pedicularis* and *Collinsia*. *Pedicularis* was excluded from the analysis because five of the six populations that utilized *Pedicularis* also utilized *Castilleja*. At population SJ, which utilized *Collinsia*, *Colacus* and *Penstemon*, *Colacus* received the majority of egg clusters in the field during the period specimens were collected for this study (Singer, M.C., unpublished data). It was therefore classified as a *Colacus* population for the purpose of the IBD, AMOVA, and genetic diversity analyses.

No significant IBD was found within the groups of populations on *Collinsia* and *Colacus*. In contrast, populations on *Castilleja* exhibited a strong IBD relationship (Table 5.9).

The current study, as well as others (Peterson 1995; Peterson 1996), suggest that IBD may be detected at some scales and not at others (see Section 5.4.4 "Differentiation at

the local scale and its correlates"). To test whether the IBD detected amongst *Castilleja* populations (average distance between populations = 352km) and not *Collinsia* populations (average distance between populations = 165km) and *Colacus* populations (average distance between populations = 162km) is due to differences in geographical scale, three *Castilleja* populations – DP, PEQ, and WH (Figure 5.1) – at the outer edge of the study range were omitted from the IBD analysis. The effects of IBD were still detected ($r = 0.42$, $P = 0.03$, Table 5.9) when the average distance between all populations was reduced to 233km.

The AMOVA analyses revealed a small but statistically significant genetic differentiation between populations utilizing *Castilleja* and *Colacus*. This was the only population group pairwise comparison that exhibited significant genetic differentiation. The other two population group pairwise comparisons – that of between *Collinsia* and *Castilleja* and between *Collinsia* and *Colacus* – did not exhibit significant genetic differentiation (Table 5.10).

Local scale. Populations in SEKI utilize *Pedicularis semibarbata* almost exclusively for oviposition. Of the eight populations included in this study, only two, RO and TJ, showed substantial use of *Castilleja applegatei* (subgenus *Castilleja*). There was no relationship between host association and genetic differentiation (Table 5.8).

Populations in YOSE use three species of the genus *Castilleja*. Populations are largely monophagous on a single species, even when another suitable species is found nearby. SI is the only population known to use two species of hosts. SL is the only population that utilizes *Castilleja miniata* (subgenus *Castilleja*), the other populations use hosts in the subgenus *Colacus*. Despite this diversity in host utilization, YOSE did not exhibit host-associated genetic structuring (Table 5.8), either at the species level, or at the sub-genus level.

5.3.4 Genetic Diversity

Genetic diversity estimates, *Hw*, at the regional and local scales were similar. The former was estimated at 0.1291 (SE(Hw) = 0.00214, Var(Hw) = 0.000005). Those of SEKI and YOSE were estimated at 0.1209 (SE(Hw) = 0.000003, Var(Hw) = 0. 000001) and 0.1254 (SE(Hw) = 0.1254, Var(Hw) = 0.000005) respectively.

The estimates of gene diversity within groups of populations on a given host were similar across the three host genera that were included in the analysis (Table 5.9). There was only one pair of gene diversity estimates that were statistically different: *Hw* for *Collinsia* populations was greater than *Hw* for *Colacus* populations (Mann Whitney, W = 138.0, P = 0.01).

The high variance among *Castilleja*-feeding larvae at TJ, seen from the scatter in the NMDS plot (Figure 5.5), suggests that the genetic diversity of individuals collected on *Castilleja* should be on average greater than the genetic diversity of individuals collected on *Pedicularis*. This proved to be true. The average heterozygosity estimates for individuals on *Castilleja* and *Pedicularis* are 0.137 (SE = 0.00731) and 0.119 (SE = 0.00699) respectively, with individuals on *Castilleja* significantly more genetically diverse than the individuals on *Pedicularis* (Mann-Whitney, W = 351616, P < 0.001).

5.4 DISCUSSION

There have been numerous studies investigating IBD to date (reviewed in Peterson and Denno 1998). There have also been studies investigating host-associated genetic differentiation (reviewed in Chapter 3), but to date, there has been no investigation into the relative effects of spatial and ecological barriers on the genetic structure of natural insect populations. The sedentary nature and patchy distribution of *E. editha* populations (Ehrlich 1965; Harrison 1989), combined with known spatial variation in their use of hosts (Singer 1994; Singer and Thomas 1996) makes this an ideal system for studies on

relationships between population structure, resource adaptation, and genetic differentiation.

In their review of how diet breadth affects IBD in phytophagous insects, Peterson and Denno (1998) tested the hypothesis that IBD increases with decreasing diet breadth. The present study looks at a related but different question: to what degree is genetic differentiation between populations associated with the differences between them in host use after taking into account the effects of spatial isolation?

5.4.1 Population Differentiation

Wright (1978) suggested that F_{ST} values of 0.05 to 0.15 indicate moderate genetic differentiation, while those of 0.15 to 0.25 indicate high genetic differentiation. At the regional scale, *E. editha* populations separated by an average of 259km demonstrated moderately high differentiation with a fixation index of $\Phi_{ST} = 0.19$ (Table 5.7). A similar estimate was obtained for a related European species, *Euphydryas aurinia*, from larvae collected on two hosts ($\Phi_{ST} = 0.18$, $P < 0.001$, average distance between populations = 69km, number of populations = 11) (see Chapter 3 "Host associated genetic differentiation in allopatric populations of the marsh fritillary *Euphydryas aurinia"*). This fixation index contrasts with estimates in Slatkin (1987) which analyzed allozyme data from 21 *E. editha* populations used in the McKechnie et al. (1975) study. The north-south extent of these populations in the Slatkin (1987) study is comparable to the one used for this study. F_{ST} estimates for 7 of the 8 allozyme loci were below 0.052, for an average F_{ST} of 0.04 (Slatkin 1987). On the other hand, Britten *et. al.*'s (1995) allozyme study of *E. editha* populations in the Great Basin and central Rocky Mountains indicated an overall F_{ST} of 0.209 across a geographical range from western Nevada to central Colorado. This estimate is closer to the one estimated in the current study, but over a wider geographical range.

Part of the reason for this discrepancy may be attributed to the more limited genetic resolution afforded by allozymes. Allozymes, although offering the advantages of co-dominance, are unlikely to show high levels of variation, as shown for other butterflies (Saccheri et al. 1999). This may lead to an underestimate of the level of differentiation. Allozyme markers have been also shown under certain cases to be under the influence of natural selection (Anderson and Oakeshott 1984; Begun and Aquadro 1994; Hedrick 1999) and thus not suitable as neutral molecular markers. AFLP studies typically employ a large number of markers. These markers are randomly scattered throughout the genome, which helps to decrease the chances that a high proportion of markers are under selection.

Previous *E. editha* studies (Baughman et al. 1967; Slatkin 1987) have assumed that the low genetic differentiation detected in their studies reflected either extensive gene flow or strong stabilizing selection throughout the survey range. Extensive current gene flow in *E. editha* is unlikely, and Baughman et al. (1967) suggested that this pattern was due to historical patterns of gene flow (Slatkin 1987) and not necessarily the result of contemporary processes. Due to shared ancestry, population genetic analysis yields a small but significant F_{ST}, incorrectly implying some amount of gene flow between the subpopulations when there may be none.

Shared ancestry is almost certain to be a confounding factor in any estimates of current gene flow drawn from this study (see below, Section 5.4.3, "IBD within populations grouped by host"). At the regional scale, correlation of genetic differentiation to geographical distance was highly significant ($r = 0.55$, $P < 0.001$), suggesting effects of IBD. Given the high fixation index between populations ($\Phi_{ST} = 0.19$, $P < 0.001$) and the geographical distances between populations (mean distance = 259km), a component of this correlation is likely to be due to historical gene flow. Contemporary gene flow may be better estimated at the local scales using the two replicates of local scale population groups in this study (see Section 5.4.4 "Differentiation at the local scale and its correlates")

5.4.2 Host-Associated Differentiation

In addition to structuring by geographical distance, both the NMDS plot (Figure 5.2) and the UPGMA tree (Figure 5.3) suggest genetic structuring mediated by host association. The effects of host association, assessed using a Mantel test between the host association matrices and the genetic differentiation matrix, were highly significant when host plants were classified at the species and sub-genus level (Table 5.8). This result is in direct contrast to the results of Baughman et al. (1990) who suggested that ecological differences between *E. editha* populations, such as oviposition host choice, are not important correlates of phylogenetic divergence.

Despite the moderately strong effects of host association detected at the species and sub-genus level using Mantel tests, the AMOVA analysis revealed only a small but statistically significant genetic differentiation between populations on *Castilleja* and populations on *Colacus* (F_{CT} = 0.02, P = 0.03). At sites where species of both sub-genera occur in sympatry, *E. editha* populations have used one but not both. This may explain the small but detectable amount of genetic differentiation found between these two population groups. None of the other population group pairs showed any detectable genetic differentiation, although there is a slight but statistically insignificant trend for populations on *Collinsia* to be differentiated from those on *Castilleja* (F_{CT} = 0.01, P = 0.07).

These AMOVA results indicate that the degree of host-associated genetic differentiation is less than that suggested by the Mantel tests. One possible reason for this disparity is the reduction in statistical power in the AMOVA analysis due to the exclusion of *Pedicularis*, *Penstemon*, and *Plantago* populations.

Within each group of population on any given host, fixation indices between populations ranged from Φ_{ST} = 0.17 (*Colacus*) to Φ_{ST} = 0.20 (*Castilleja*) (Table 5.9). Five of the six pairwise comparisons of genetic diversities between groups of populations on any two

given hosts showed no detectable difference. These results suggest that similar demographic processes are acting on all populations, regardless of host association.

5.4.3 IBD Within Populations Grouped by Host

Radtkey & Singer's (1995) mtDNA analysis suggested the subgenus *Castilleja* as the ancestral host plant of *E. editha*. Populations utilizing *Castilleja* appear more widely-dispersed in the NMDS plot than populations on other hosts (Figure 5.2). This contrasts with populations utilizing *Collinsia* which are clustered around the origin of the NMDS plot. Populations east of the Sierra Nevada are not known to utilize *Collinsia*. While these results suggest that populations utilizing more recently colonized hosts tend to cluster around the origin of the NMDS plot, there is insufficient evidence to suggest that this pattern holds true in general. The number of populations utilizing *Pedicularis*, *Penstemon*, and *Plantago* in this study is too small to make any further generalizations between host use novelty and the distribution pattern of populations on the NMDS plot. However, such patterns observed in an NMDS plot may provide an incentive to conduct an IBD analysis of populations grouped by their principle host. This is discussed next.

Strong evidence for IBD was found among populations associated with *Castilleja* (Table 5.9). Part of this highly significant correlation between genetic differentiation and geographic distance ($r = 0.7$, $P < 0.001$) is likely to be attributed to historical gene flow. This is likely attributed to the range expansion of *E. editha* from Colorado to the Sierra Nevada. The high fixation index between populations ($\Phi_{ST} = 0.2$, $P < 0.001$) and the geographical scale (average distance between populations = 352km with PEQ, 281km without) argue against high levels of contemporary gene flow.

Collinsia populations show no evidence of IBD. One possible explanation for this pattern is to postulate multiple colonizations of *Collinsia*. Each colonization constitutes a founding event which eliminates the IBD relationship through a random sampling of alleles that would have otherwise maintained the IBD 'signature' resulting from the range

expansion from Colorado to the Sierra Nevada. An alternate hypothesis would be to assume that *Collinsia* was colonized once, followed by a spread of *Collinsia* feeding populations through the landscape. However, this is not likely for two reasons. First, it is likely that the subsequent expansion of *Collinsia* populations would result in detectable IBD. Second, if *Collinsia* populations arose from a single founding event, one may expect that *Collinsia* populations have lower genetic diversity when compared to populations utilizing other hosts. However, as reported earlier (see Section 5.3.4 "Genetic diversity"), the estimates of gene diversity within groups of populations on a given host were similar across the three host genera that were included in the analysis (Table 5.9). There was only one genetic diversity estimate (*Hw*) comparison that was statistically significant: *Hw* for *Collinsia* populations was greater than, and not less than, *Hw* for *Colacus* populations. These observations suggest that the "multiple *Collinsia* colonization" hypothesis is more likely to account for the lack of detectable IBD between *Collinsia* populations.

Similar to *Collinsia*, populations on *Colacus* exhibit a moderately high fixation index ($\Phi_{ST} = 0.15$, $P < 0.001$). These populations also showed no evidence of IBD at the regional scale. This is in direct contradiction to the strong effect of distance on genetic differentiation observed at the local scale at YOSE ($r = 0.66$, $P < 0.001$) where seven out of the eight populations utilize *Colacus*. Recall that at the regional level, the significant relationship between genetic differentiation and host association was lost when *Castilleja* and *Colacus* were amalgamated into the genus *Castilleja*. There was also a small but statistically significant genetic differentiation between populations on *Castilleja* and populations on *Colacus* (Table 5.10). This, together with the observation of the lack of IBD among *Colacus* populations but strong IBD among *Castilleja* populations, is further reason to believe that *Castilleja* and *Colacus* have different effects on genetic differentiation and should therefore be analyzed separately, even if they do both belong to the genus *Castilleja*. One possible explanation for the lack of IBD among *Colacus*

populations is to postulate that the population genetic structure for this group of populations is determined more by selection or drift than by migration.

5.4.4 Differentiation at the Local Scale

Wright (1978) notes that even though F_{ST}s of less than 0.05 indicate little genetic differentiation, they are by no means negligible (Hartl and Clark 1997). The Φ_{ST} estimates of both SEKI and YOSE fall into this range, reflecting a relatively low but still significant degree of differentiation with a fixation index of 0.03 (Table 5.7). There is a strong, significant effect of distance on the genetic structure of populations in YOSE ($r = 0.66$, $P < 0.001$), while SEKI shows a much weaker trend at the same geographical scale ($r = 0.24$, $p = 0.25$). The significance of this difference in IBD between SEKI and YOSE has not yet been tested. This is due to the current lack of a readily available statistical package to assess the statistical significance of the difference of any two given Pearson correlation coefficients calculated by a permutation test.

A lack of IBD at SEKI could imply either that there is very limited gene flow between populations which are then subject to the effects of drift, or that populations are in panmixia with sufficient gene flow to homogenize the effects of local selection and drift. Given the low degree of genetic differentiation at this scale and the similarity in habitat between populations, the latter is more likely. There is ample evidence of metapopulation effects at SEKI (Singer and Thomas 1996; Thomas et al. 1996; Boughton 1999; Boughton 2000), which suggests that the panmixia in SEKI is likely to arise from stepping-stone roles played by multiple small populations dotted throughout the landscape, most of which were not sampled in this study. Moreover, a previous mark-recapture study (Singer MC, Gilbert L., unpublished) found that adult insects at SEKI are less sedentary than insects at YOSE. This may help explain the why IBD was found at YOSE, but not at SEKI.

Populations in YOSE use two *Colacus* species and one *Castilleja*. No significant host-associated differentiation was found, but the power to find it was low because of the low number of populations on one host.

Populations in SEKI utilize *Pedicularis* and *Castilleja*. Genetic differentiation between SEKI populations was neither associated with distance, elevation, nor host utilization (Tables 5.7, 5.8). The latter result is reinforced by an AMOVA analysis, which did not reveal detectable differentiation between populations on the two hosts (F_{CT} = -0.004, P = 0.53). The power to detect such differentiation is low, granted that there is only one population on *Castilleja*. Despite this failure to find an expected host effect, an unexpected host effect was detected at SEKI. One of the SEKI populations using two hosts, TJ, showed a difference between individuals sampled from two hosts: genetic variance among individuals was higher on *Castilleja* than on *Pedicularis* (Figure 5.5). One possible mechanism for this pattern is stabilizing selection on *Pedicularis*.

Table 5.1: Populations included in the study.

Population Code	Population Name	Year(s) Collected	Sample Size	Genus: Collinsia			Genus: Castilleja colacus			Genus: Castilleja castilleja			Genus: Pedicularis		Genus: Plantago		Penstemon rydbergii
				Collinsia Tinctoria	Collinsia parviflora	Collinsia torreyi	Castilleja nana	Castilleja lemmoni	Castilleja densiflorus	Castilleja foliolosa	Castilleja miniata	Castilleja applegatei	Pedicularis semibarbata	Pedicularis densiflora	Plantago erecta	Plantago lanceolata	
AFL	Agua Fria	1999 – 2000	10	1													
AU	Gold Lake	1999 – 2000	11								1						
BB (SEKI)	Big Baldy	1999 – 2001	9										1				
BF	Bircham Flat	1999 – 2000	10		1												
BLC	Balch Camp	1996	2	1													
BM	Big Meadow	1996	14									1	1				
BR (SEKI)	Buck Rock	2000 – 2001	10										1				
CG	Coarse Gold	2002	3	1													
CM (SEKI)	Colony Meadows	2000	12										1				
DM (YOSE)	Dana Meadows	1999 – 2000	14					1									
DP	Del Peurto Canyon	2002	13							1				1			
EP	Ebbets' Pass	1999 – 2000	10				1										
FR	Frenchman Lake	1999 – 2000	15				1										1
GDL (YOSE)	Gardisky Lake	1999	7				1										
GLA	Glen Alpine	2003	3								1	1	1				
IF	Indian Flat	1999 – 2001	10	1													
IOH	Iowa Hill	2003	4	1													
LBO (SEKI)	Lower Buck Rock	2002	6										1				
LK	Leek Springs	1999	11			1											
LOP (SEKI)	Look Out Point	2003	5										1				
MD (YOSE)	Mount Dana	2000	10				1										
MG	McGee	1999 – 2000	39		1												1
MH	Morgan Hill	1996	25						1						1		

Table 5.1 (Continued)

Population Code	Population Name	Year(s) Collected	Sample Size	Genus: Collinsia			Genus: Castilleja colacus			Genus: Castilleja castilleja			Genus: Pedicu-laris		Genus: Plantago		Penstemon rydbergii
				Collinsia Tinctoria	Collinsia parviflora	Collinsia torreyi	Castilleja nana	Castilleja lemmoni	Castilleja densiflorus	Castilleja foliolosa	Castilleja miniata	Castilleja applegatei	Pedicularis semibarbata	Pedicularis densiflora	Plantago erecta	Plantago lanceolata	
MUD	Mud Creek	2003	6	1													
PEQ	Pequop Mountain	1990	5									1					
PI	Piute	2002	9									1	1				
PP (YOSE)	Parker Pass	1999 – 2000	10				1										
RM (SEKI)	Rabbit Meadow	1990, 1999, 2001 – 2003	22										1				
RO (SEKI)	Rowell Meadows	2000	10									1					
SI (YOSE)	Spillway Lake	1999 – 2000	10				1	1									
SJ	Sonora Junction	1999 – 2001	34		1		1										1
SK	Skinner	2002	6												1		
SL (YOSE)	Saddlebag Lake	2000	9								1						
SN	Schnieder's Meadow	2002	13		1											1	1
SR (YOSE)	Saddlebag Ridge	1999 – 2000	16				1										
TAH	Tahoe Meadows	2003	4				1										
TG	Timber Gap	2000	12									1	1				
TJ (SEKI)	T Junction	1999 – 2000	32									1	1				
TM (YOSE)	Tuolumne Meadows	2000	11					1									
TR	Tamarack Ridge	1999, 2003	17			1											
WH	Wild Horse Canyon	2002	9								1						
YP	Yucca Point	2000 – 2001	10	1													

Table 5.2: Population pairwise Φ_{ST} matrix for populations at the regional scale.

	AFL	AU	BF	BLC	BM	CG	DP	EP	FR	GLA	
AFL											
AU	0.15										
BF	0.13	0.16									
BLC	0.02	0.10	0.11								
BM	0.17	0.21	0.21	0.07							
CG	0.04	0.11	0.15	0.00	0.12						
DP	0.11	0.12	0.13	0.00	0.18	0.07					
EP	0.16	0.15	0.11	0.12	0.20	0.14	0.16				
FR	0.17	0.15	0.09	0.13	0.22	0.18	0.14	0.14			
GLA	0.11	0.16	0.14	0.03	0.21	0.07	0.12	0.14	0.16		
IF	0.02	0.14	0.17	0.00	0.18	0.04	0.10	0.17	0.19	0.12	
IOH	0.07	0.11	0.12	0.00	0.17	0.04	0.05	0.12	0.11	0.08	
LK	0.10	0.14	0.13	0.04	0.18	0.06	0.09	0.14	0.15	0.11	
MG	0.28	0.29	0.27	0.28	0.29	0.30	0.27	0.28	0.25	0.29	
MH	0.10	0.16	0.13	0.02	0.17	0.07	0.08	0.17	0.18	0.13	
MUD	0.11	0.11	0.14	0.10	0.20	0.12	0.09	0.16	0.14	0.14	
PEQ	0.24	0.27	0.26	0.27	0.30	0.32	0.21	0.29	0.23	0.29	
PI	0.16	0.22	0.20	0.04	0.17	0.15	0.12	0.23	0.24	0.16	
RM	0.13	0.17	0.18	0.04	0.08	0.08	0.14	0.18	0.17	0.12	
RO	0.21	0.26	0.26	0.14	0.17	0.16	0.21	0.25	0.26	0.23	
SJ	0.14	0.15	0.05	0.12	0.21	0.14	0.14	0.10	0.08	0.13	
SK	0.18	0.27	0.26	0.10	0.23	0.19	0.22	0.27	0.27	0.20	
SL	0.16	0.12	0.12	0.15	0.17	0.18	0.14	0.13	0.14	0.19	
SN	0.16	0.17	0.08	0.15	0.23	0.18	0.15	0.13	0.10	0.17	
TAH	0.20	0.22	0.17	0.21	0.30	0.25	0.18	0.21	0.16	0.25	
TG	0.14	0.20	0.18	0.03	0.08	0.10	0.17	0.19	0.19	0.14	
TM	0.16	0.16	0.15	0.16	0.21	0.19	0.15	0.17	0.15	0.19	
TR	0.09	0.14	0.13	0.00	0.11	0.04	0.09	0.14	0.15	0.13	
WH	0.20	0.30	0.30	0.21	0.24	0.24	0.24	0.29	0.28	0.28	
YP	0.07	0.11	0.14	0.00	0.11	0.00	0.06	0.14	0.16	0.09	

MATRIX CONTINUED ON NEXT PAGE

Table 5.2 (Continued)

	IF	IOH	LK	MG	MH	MUD	PEQ	PI	RM	RO	
AFL											
AU											
BF											
BLC											
BM											
CG											
DP											
EP											
FR											
GLA											
IF											
IOH	0.08										
LK	0.11	0.07									
MG	0.30	0.29	0.28								MATRIX CONTINUED ON NEXT PAGE
MH	0.11	0.07	0.12	0.27							
MUD	0.11	0.05	0.12	0.32	0.11						
PEQ	0.24	0.29	0.25	0.40	0.23	0.26					
PI	0.14	0.18	0.15	0.34	0.15	0.18	0.29				
RM	0.15	0.14	0.17	0.24	0.16	0.16	0.21	0.17			
RO	0.21	0.23	0.25	0.34	0.23	0.26	0.32	0.26	0.08		
SJ	0.16	0.13	0.14	0.24	0.16	0.15	0.23	0.22	0.18	0.25	
SK	0.20	0.22	0.24	0.38	0.22	0.23	0.35	0.23	0.20	0.27	
SL	0.18	0.15	0.15	0.28	0.17	0.18	0.29	0.24	0.15	0.24	
SN	0.18	0.14	0.16	0.28	0.16	0.17	0.25	0.23	0.18	0.27	
TAH	0.21	0.20	0.22	0.34	0.22	0.24	0.35	0.27	0.22	0.32	
TG	0.16	0.15	0.19	0.28	0.15	0.17	0.25	0.19	0.04	0.10	
TM	0.20	0.12	0.16	0.31	0.17	0.17	0.26	0.26	0.17	0.28	
TR	0.08	0.07	0.08	0.26	0.09	0.11	0.24	0.12	0.11	0.19	
WH	0.22	0.24	0.26	0.35	0.24	0.29	0.39	0.26	0.20	0.29	
YP	0.05	0.04	0.08	0.25	0.07	0.10	0.21	0.10	0.09	0.17	

Table 5.2 (Continued)

	SJ	SK	SL	SN	TAH	TG	TM	TR	WH	YP
AFL										
AU										
BF										
BLC										
BM										
CG										
DP										
EP										
FR										
GLA										
IF										
IOH										
LK										
MG										
MH										
MUD										
PEQ										
PI										
RM										
RO										
SJ										
SK	0.25									
SL	0.11	0.29								
SN	0.10	0.27	0.18							
TAH	0.15	0.31	0.24	0.18						
TG	0.18	0.19	0.17	0.20	0.25					
TM	0.13	0.29	0.09	0.17	0.23	0.19				
TR	0.14	0.20	0.14	0.15	0.20	0.12	0.16			
WH	0.28	0.27	0.32	0.31	0.37	0.23	0.30	0.21		
YP	0.15	0.15	0.13	0.14	0.18	0.11	0.15	0.04	0.18	

Table 5.3: Population pairwise Φ_{ST} matrix for SEKI (lower triangular) and YOSE (upper triangular).

		YOSE								
		DM	**GDL**	**MD**	**PP**	**SI**	**SL**	**SR**	**TM**	
SEKI	**BB**		**0.025**	**0.024**	0.020	0.009	**0.029**	**0.014**	**0.064**	**TM**
	BR	0.018		0.001	0.037	**0.025**	0.020	0.009	**0.067**	**SR**
	CM	0.001	**0.032**		0.009	0.001	**0.021**	0.012	**0.050**	**SL**
	LBO	0.022	0.015	0.027		0.000	**0.043**	**0.035**	**0.080**	**SI**
	LOP	**0.051**	0.017	**0.048**	**0.071**		**0.037**	**0.014**	**0.063**	**PP**
	RM	**0.019**	0.013	**0.016**	**0.027**	0.022		0.004	**0.089**	**MD**
	RO	**0.067**	**0.064**	**0.066**	**0.067**	**0.096**	**0.076**		**0.057**	**GDL**
	TJ	**0.014**	**0.023**	**0.015**	0.013	**0.031**	**0.017**	**0.068**		**DM**
		TJ	**RO**	**RM**	**LOP**	**LBO**	**CM**	**BR**	**BB**	

Table 5.4: First ten eigenvalues from NMDS analysis of Nei's corrected average pairwise differences for populations at the regional scale.

Dimension Number	Eigenvalue	Individual Percent	Cumulative Percent	Cumulative Bar Chart
1	1354.48	25.93	25.93	□□□□□□□□□□□□□□
2	679.34	13.01	38.94	□□□□□□□□
3	595.02	11.39	50.33	□□□□□□□
4	369.85	7.08	57.41	□□□□□
5	336.60	6.44	63.85	□□□□
6	255.94	4.90	68.75	□□□
7	242.71	4.65	73.40	□□□
8	152.67	2.92	76.32	□□
9	135.29	2.59	78.91	□□
10	101.90	1.95	80.86	□□

Table 5.5: Population pairwise distances in km (lower triangular) and elevation differences in meters (upper triangular).

	AFL	AU	BF	BLC	BM	CG	DP	EP	FR	GLA	
AFL		1364.0	1311.0	130.0	1720.0	70.0	130.0	2118.0	1076.0	1658.0	
AU	247.4		53.0	1494.0	356.0	1294.0	1494.0	754.0	288.0	294.0	
BF	117.9	171.0		1441.0	409.0	1241.0	1441.0	807.0	235.0	347.0	
BLC	104.5	336.0	175.2		1850.0	200.0	0.0	2248.0	1206.0	1788.0	
BM	233.7	467.0	301.5	132.1		1650.0	1850.0	398.0	644.0	62.0	
CG	38.8	280.1	134.4	65.8	195.3		200.0	2048.0	1006.0	1588.0	
DP	123.1	259.5	208.4	211.6	322.4	152.4		2248.0	1206.0	1788.0	
EP	118.3	144.2	34.1	193.4	323.3	143.3	189.0		1042.0	460.0	
FR	272.6	49.7	178.4	351.8	479.8	301.5	301.0	158.9		582.0	
GLA	153.7	99.9	74.3	236.5	367.2	183.1	199.8	44.4	118.9		
IF	25.3	233.3	93.8	106.9	238.8	46.8	142.1	97.8	255.1	136.5	
IOH	194.6	67.1	147.1	291.5	423.6	230.5	192.8	113.9	112.5	75.1	
LK	128.8	119.4	72.5	217.7	349.5	160.7	171.6	38.4	144.7	28.7	
MG	127.1	317.4	147.0	64.1	159.6	97.0	248.5	174.0	324.4	218.2	
MH	136.4	283.1	229.1	216.8	320.6	161.6	23.9	211.1	324.8	223.3	
MUD	306.0	94.5	254.6	404.5	536.6	342.8	279.5	223.3	129.2	180.3	
PEQ	616.0	540.3	509.2	608.7	663.6	613.3	717.3	529.7	493.8	531.8	
PI	268.3	507.3	344.2	171.3	46.2	230.9	345.8	364.5	522.1	407.8	
RM	134.5	363.7	200.0	30.0	103.3	95.8	238.7	220.4	377.8	264.0	
RO	142.1	367.6	202.1	37.8	99.6	103.5	248.2	223.8	380.3	267.7	
SJ	107.7	178.3	10.6	165.3	292.4	123.8	201.2	37.0	187.5	80.0	
SK	509.5	746.7	580.2	411.3	279.7	472.1	578.6	603.0	758.6	646.9	
SL	84.4	222.2	54.4	121.1	247.3	87.8	197.9	78.4	232.6	122.7	
SN	179.8	93.7	80.7	253.8	381.8	205.7	233.6	62.3	98.1	33.9	
TAH	201.6	76.2	103.1	276.6	404.4	228.2	248.9	84.8	75.4	50.7	
TG	170.0	397.8	232.0	65.7	69.5	131.2	270.8	254.0	410.3	298.0	
TM	72.1	228.6	64.4	110.9	239.1	74.7	188.2	84.5	241.2	128.7	
TR	85.7	304.8	141.9	33.7	162.2	51.0	203.4	161.4	319.0	205.0	
WH	469.0	698.5	529.3	367.0	235.4	430.7	549.9	554.3	706.9	598.6	
YP	125.7	351.8	187.4	22.1	115.2	87.2	233.6	208.2	365.3	252.0	

MATRIX CONTINUED ON NEXT PAGE

Table 5.5 (Continued)

	IF	IOH	LK	MG	MH	MUD	PEQ	PI	RM	RO	
AFL	80.0	130.0	2110.0	1700.0	230.0	120.0	1370.0	1770.0	1737.0	1948.0	
AU	1444.0	1494.0	746.0	336.0	1594.0	1244.0	6.0	406.0	373.0	584.0	
BF	1391.0	1441.0	799.0	389.0	1541.0	1191.0	59.0	459.0	426.0	637.0	
BLC	50.0	0.0	2240.0	1830.0	100.0	250.0	1500.0	1900.0	1867.0	2078.0	
BM	1800.0	1850.0	390.0	20.0	1950.0	1600.0	350.0	50.0	17.0	228.0	
CG	150.0	200.0	2040.0	1630.0	300.0	50.0	1300.0	1700.0	1667.0	1878.0	
DP	50.0	0.0	2240.0	1830.0	100.0	250.0	1500.0	1900.0	1867.0	2078.0	
EP	2198.0	2248.0	8.0	418.0	2348.0	1998.0	748.0	348.0	381.0	170.0	
FR	1156.0	1206.0	1034.0	624.0	1306.0	956.0	294.0	694.0	661.0	872.0	
GLA	1738.0	1788.0	452.0	42.0	1888.0	1538.0	288.0	112.0	79.0	290.0	
IF		50.0	2190.0	1780.0	150.0	200.0	1450.0	1850.0	1817.0	2028.0	
IOH	185.0		2240.0	1830.0	100.0	250.0	1500.0	1900.0	1867.0	2078.0	
LK	113.9	77.9		410.0	2340.0	1990.0	740.0	340.0	373.0	162.0	MATRIX
MG	115.8	284.0	206.3		1930.0	1580.0	330.0	70.0	37.0	248.0	CONTINUED
MH	157.5	216.2	195.0	258.6		350.0	1600.0	2000.0	1967.0	2178.0	ON
MUD	297.9	113.2	190.0	396.1	300.5		1250.0	1650.0	1617.0	1828.0	NEXT
PEQ	591.0	584.5	555.2	545.0	738.3	619.4		400.0	367.0	578.0	PAGE
PI	276.2	461.2	388.8	204.7	341.1	573.7	706.4		33.0	178.0	
RM	136.5	320.7	246.2	70.2	242.4	433.8	611.8	144.3		211.0	
RO	142.7	325.9	250.7	66.8	252.3	439.1	605.2	142.2	10.3		
SJ	83.4	150.9	74.4	139.1	221.5	259.9	517.0	334.7	190.5	192.9	
SK	517.2	702.2	629.0	435.0	570.4	814.9	858.7	241.2	383.0	379.2	
SL	59.6	189.3	111.3	95.6	215.1	300.8	531.7	289.7	145.5	147.8	
SN	160.0	93.3	62.0	227.5	257.0	184.1	500.0	424.0	279.7	282.3	
TAH	182.3	91.0	79.4	249.5	272.7	169.5	494.8	446.7	302.4	304.9	
TG	172.1	355.9	280.9	92.4	272.9	469.1	619.6	113.0	35.7	30.3	
TM	47.8	192.2	114.6	91.8	204.6	304.6	544.4	280.9	136.5	139.5	
TR	81.3	262.9	187.6	47.9	212.2	376.1	583.8	203.2	59.0	63.2	
WH	473.8	658.1	582.8	382.5	544.4	771.3	787.3	204.3	337.5	332.2	
YP	126.1	309.5	234.6	58.9	238.5	422.7	602.4	156.8	12.7	16.6	

Table 5.5 (Continued)

	SJ	SK	SL	SN	TAH	TG	TM	TR	WH	YP
AFL	1520.0	130.0	2481.0	1070.0	1370.0	2074.0	1990.0	2020.0	2070.0	588.0
AU	156.0	1494.0	1117.0	294.0	6.0	710.0	626.0	656.0	706.0	776.0
BF	209.0	1441.0	1170.0	241.0	59.0	763.0	679.0	709.0	759.0	723.0
BLC	1650.0	0.0	2611.0	1200.0	1500.0	2204.0	2120.0	2150.0	2200.0	718.0
BM	200.0	1850.0	761.0	650.0	350.0	354.0	270.0	300.0	350.0	1132.0
CG	1450.0	200.0	2411.0	1000.0	1300.0	2004.0	1920.0	1950.0	2000.0	518.0
DP	1650.0	0.0	2611.0	1200.0	1500.0	2204.0	2120.0	2150.0	2200.0	718.0
EP	598.0	2248.0	363.0	1048.0	748.0	44.0	128.0	98.0	48.0	1530.0
FR	444.0	1206.0	1405.0	6.0	294.0	998.0	914.0	944.0	994.0	488.0
GLA	138.0	1788.0	823.0	588.0	288.0	416.0	332.0	362.0	412.0	1070.0
IF	1600.0	50.0	2561.0	1150.0	1450.0	2154.0	2070.0	2100.0	2150.0	668.0
IOH	1650.0	0.0	2611.0	1200.0	1500.0	2204.0	2120.0	2150.0	2200.0	718.0
LK	590.0	2240.0	371.0	1040.0	740.0	36.0	120.0	90.0	40.0	1522.0
MG	180.0	1830.0	781.0	630.0	330.0	374.0	290.0	320.0	370.0	1112.0
MH	1750.0	100.0	2711.0	1300.0	1600.0	2304.0	2220.0	2250.0	2300.0	818.0
MUD	1400.0	250.0	2361.0	950.0	1250.0	1954.0	1870.0	1900.0	1950.0	468.0
PEQ	150.0	1500.0	1111.0	300.0	0.0	704.0	620.0	650.0	700.0	782.0
PI	250.0	1900.0	711.0	700.0	400.0	304.0	220.0	250.0	300.0	1182.0
RM	217.0	1867.0	744.0	667.0	367.0	337.0	253.0	283.0	333.0	1149.0
RO	428.0	2078.0	533.0	878.0	578.0	126.0	42.0	72.0	122.0	1360.0
SJ		1650.0	961.0	450.0	150.0	554.0	470.0	500.0	550.0	932.0
SK	571.4		2611.0	1200.0	1500.0	2204.0	2120.0	2150.0	2200.0	718.0
SL	45.1	526.3		1411.0	1111.0	407.0	491.0	461.0	411.0	1893.0
SN	89.5	660.8	134.6		300.0	1004.0	920.0	950.0	1000.0	482.0
TAH	112.1	683.3	157.2	22.8		704.0	620.0	650.0	700.0	782.0
TG	222.8	349.0	177.7	312.3	334.9		84.0	54.0	4.0	1486.0
TM	54.4	518.6	13.2	143.2	165.9	169.7		30.0	80.0	1402.0
TR	132.1	441.9	87.6	221.0	243.7	93.4	77.9		50.0	1432.0
WH	521.1	72.0	476.4	609.9	632.0	302.2	469.9	395.4		1482.0
YP	178.0	394.9	133.0	267.3	290.0	46.4	124.2	47.1	348.6	

Table 5.6: Host association distance matrices by species (lower triangular) and by sub-genera (upper triangular).

	AFL	AU	BF	BLC	BM	CG	DP	EP	FR	GLA	
AFL		1.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	
AU	1.00		1.00	1.00	0.33	1.00	0.33	1.00	1.00	0.33	
BF	1.00	1.00		0.00	1.00	0.00	1.00	1.00	1.00	1.00	
BLC	0.00	1.00	1.00		1.00	0.00	1.00	1.00	1.00	1.00	
BM	1.00	1.00	1.00	1.00		1.00	0.00	1.00	1.00	0.00	
CG	0.00	1.00	1.00	0.00	1.00		1.00	1.00	1.00	1.00	
DP	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	0.00	
EP	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.33	1.00	
FR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33		1.00	
GLA	1.00	0.50	1.00	1.00	0.20	1.00	1.00	1.00	1.00		
IF	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	
IOH	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	
LK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MG	1.00	1.00	0.33	1.00	1.00	1.00	1.00	1.00	0.50	1.00	
MH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MUD	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	
PEQ	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00	0.50	
PI	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.20	
RM	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00	0.50	
RO	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00	0.50	
SJ	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50	0.20	1.00	
SK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
SL	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	
SN	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	0.60	1.00	
TAH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.33	1.00	
TG	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.20	
TM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
TR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
WH	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	
YP	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	

MATRIX CONTINUED ON NEXT PAGE

Table 5.6 (Continued)

	IF	IOH	LK	MG	MH	MUD	PEQ	PI	RM	RO	
AFL	0.00	0.00	0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
AU	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.33	1.00	0.00	
BF	0.00	0.00	0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
BLC	0.00	0.00	0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
BM	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.00	0.33	0.33	
CG	0.00	0.00	0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
DP	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.00	0.33	0.33	
EP	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00	1.00	
FR	1.00	1.00	1.00	0.50	0.50	1.00	1.00	1.00	1.00	1.00	
GLA	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.00	0.33	0.33	
IF		0.00	0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
IOH	0.00		0.00	0.33	1.00	0.00	1.00	1.00	1.00	1.00	
LK	1.00	1.00		0.33	1.00	0.00	1.00	1.00	1.00	1.00	MATRIX
MG	1.00	1.00	1.00		1.00	0.33	1.00	1.00	1.00	1.00	CONTINUED
MH	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	ON
MUD	0.00	0.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	NEXT
PEQ	1.00	1.00	1.00	1.00	1.00	1.00		0.33	1.00	0.00	PAGE
PI	1.00	1.00	1.00	1.00	1.00	1.00	0.33		0.33	0.33	
RM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33		1.00	
RO	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.33	1.00		
SJ	1.00	1.00	1.00	0.20	1.00	1.00	1.00	1.00	1.00	1.00	
SK	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00	1.00	
SL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
SN	1.00	1.00	1.00	0.20	1.00	1.00	1.00	1.00	1.00	1.00	
TAH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
TG	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.00	0.33	0.33	
TM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
TR	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
WH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
YP	0.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	

Table 5.6 (Continued)

	SJ	SK	SL	SN	TAH	TG	TM	TR	WH	YP
AFL	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
AU	1.00	1.00	0.00	1.00	1.00	0.33	1.00	1.00	0.00	1.00
BF	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
BLC	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
BM	1.00	1.00	0.33	1.00	1.00	0.00	1.00	1.00	0.33	1.00
CG	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
DP	1.00	1.00	0.33	1.00	1.00	0.00	1.00	1.00	0.33	1.00
EP	0.50	1.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00
FR	0.20	1.00	1.00	0.60	0.33	1.00	0.33	1.00	1.00	1.00
GLA	1.00	1.00	0.33	1.00	1.00	0.00	1.00	1.00	0.33	1.00
IF	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
IOH	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
LK	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
MG	0.20	1.00	1.00	0.20	1.00	1.00	1.00	0.33	1.00	0.33
MH	0.60	0.33	1.00	0.60	0.33	1.00	0.33	1.00	1.00	1.00
MUD	0.50	1.00	1.00	0.50	1.00	1.00	1.00	0.00	1.00	0.00
PEQ	1.00	1.00	0.00	1.00	1.00	0.33	1.00	1.00	0.00	1.00
PI	1.00	1.00	0.33	1.00	1.00	0.00	1.00	1.00	0.33	1.00
RM	1.00	1.00	1.00	1.00	1.00	0.33	1.00	1.00	1.00	1.00
RO	1.00	1.00	0.00	1.00	1.00	0.33	1.00	1.00	0.00	1.00
SJ		1.00	1.00	0.33	0.50	1.00	0.50	0.50	1.00	0.50
SK	1.00		1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00
SL	1.00	1.00		1.00	1.00	0.33	1.00	1.00	0.00	1.00
SN	0.33	1.00	1.00		1.00	1.00	1.00	0.50	1.00	0.50
TAH	0.50	1.00	1.00	1.00		1.00	0.00	1.00	1.00	1.00
TG	1.00	1.00	1.00	1.00	1.00		1.00	1.00	0.33	1.00
TM	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00
TR	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00	0.00
WH	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00		1.00
YP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 5.7: Correlation between Φ_{ST} and geographical distance and elevation.

Spatial Scale	Num. of Pops	Φ_{ST}	Geographical Distance				Elevation	
			Average Dist. (km)	Φ_{ST} x Dist.	(Φ_{ST} x Dist.) with host held constant		Average Elev. (m)	Φ_{ST} x Elevation
					Species Level	Subgenus Level		
SEKI	8	**0.03**	8.7	0.24 (P = 0.25)	N.A.	N.A.	115	-0.02 (P = 0.51)
YOSE	8	**0.03**	9.3	**0.66** (P < 0.001)	**0.67** (P < 0.001)	**0.66** (P < 0.001)	367	0.49 (P = 0.1)
Regional	30	**0.19**	259	**0.55** (P < 0.001)	**0.56** (P < 0.001)	**0.55** (P < 0.001)	1013	-0.04 (P = 0.27)

Table 5.8: Correlation between Φ_{ST} and host association at the species, sub-genus, and genus levels.

Spatial Scale	Φ_{ST}	Species Level			Sub-genus Level			Genus Level		
		Num. of Species	Φ_{ST} x Host	(Φ_{ST} x Host) with distance held constant	Num. of Sub Genera	Φ_{ST} x Host	(Φ_{ST} x Host) with distance held constant	Num. of Genera	Φ_{ST} x Host	(Φ_{ST} x Host) with distance held constant
SEKI	0.03	---			---			2	0.71 (P = 0.11)	0.73 (P = 0.09)
YOSE	0.03	3	0.32 (P = 0.14)	0.33 (P = 0.13)	2	0.10 (P = 0.25)	0.05 (P = 0.34)			
Regional	0.19	14	**0.238** (P = 0.0001)	**0.24** (P = 0.0001)	6	**0.21** (P = 0.0004)	**0.17** (P = 0.003)	5	0.08 (P = 0.08)	0.05 (P = 0.16)

Table 5.9: IBD analysis and genetic diversity estimates of populations grouped by host utilization.

Host Genus	Num. of Pops.	Φ_{ST}	Av. Dist. between pops (km)	Φ_{ST} x Distance	Genetic diversity estimates		
					Population	Population *H*, averaged over all loci	*Hw* (SE), averaged over *H*
Collinsia	12	0.18	165	0.16 (P = 0.15)	AFL	0.1373	0.1296 (0.00440)
					BF	0.1241	
					BLC	0.1491	
					CG	0.1431	
					IF	0.1429	
					IOH	0.1389	
					LK	0.1315	
					MG	0.0957	
					MUD	0.1228	
					SN	0.1107	
					TR	0.1245	
					YP	0.1339	
Colacus	6	0.17	162	0.28 (P = 0.25)	EP	0.1219	0.1196 (0.00072)
					FR	0.1170	
					MH	0.1180	
					SJ	0.1201	
					TAH	0.1204	
					TM	0.1198	
Castilleja (All populations)	10	0.20	352	**0.70** (P < 0.001)	AU	0.1290	0.1230 (0.00350)
					BM	0.1241	
					DP	0.1306	
					GLA	0.1407	
					PEQ	0.1041	
					PI	0.1160	
					RO	0.1299	
					SL	0.1173	
					TG	0.1284	
					WH	0.1095	

Table 5.9 (Continued)

Host Genus	Num. of Pops.	Φ_{ST}	Av. Dist. between pops (km)	Φ_{ST} x Distance	Genetic diversity estimates		
					Population	Population *H*, averaged over all loci	*Hw* (SE), averaged over *H*
Castilleja (Omitted: DP, PEQ, WH)	7	0.18	233	**0.42** (P = 0.03)	AU	0.1290	0.1265 (0.00317)
					BM	0.1241	
					GLA	0.1407	
					PI	0.1160	
					RO	0.1299	
					SL	0.1173	
					TG	0.1284	

Table 5.10: Pairwise AMOVA F_{CT} coefficients from analysis of populations grouped by host.

	Collinsia	**Castilleja**	**Colacus**
Collinsia	--	--	--
Castilleja	0.01 (P = 0.07)	--	--
Colacus	0 (P = 0.41)	**0.02** (P = 0.03)	--

Figure 5.1: Study sites.

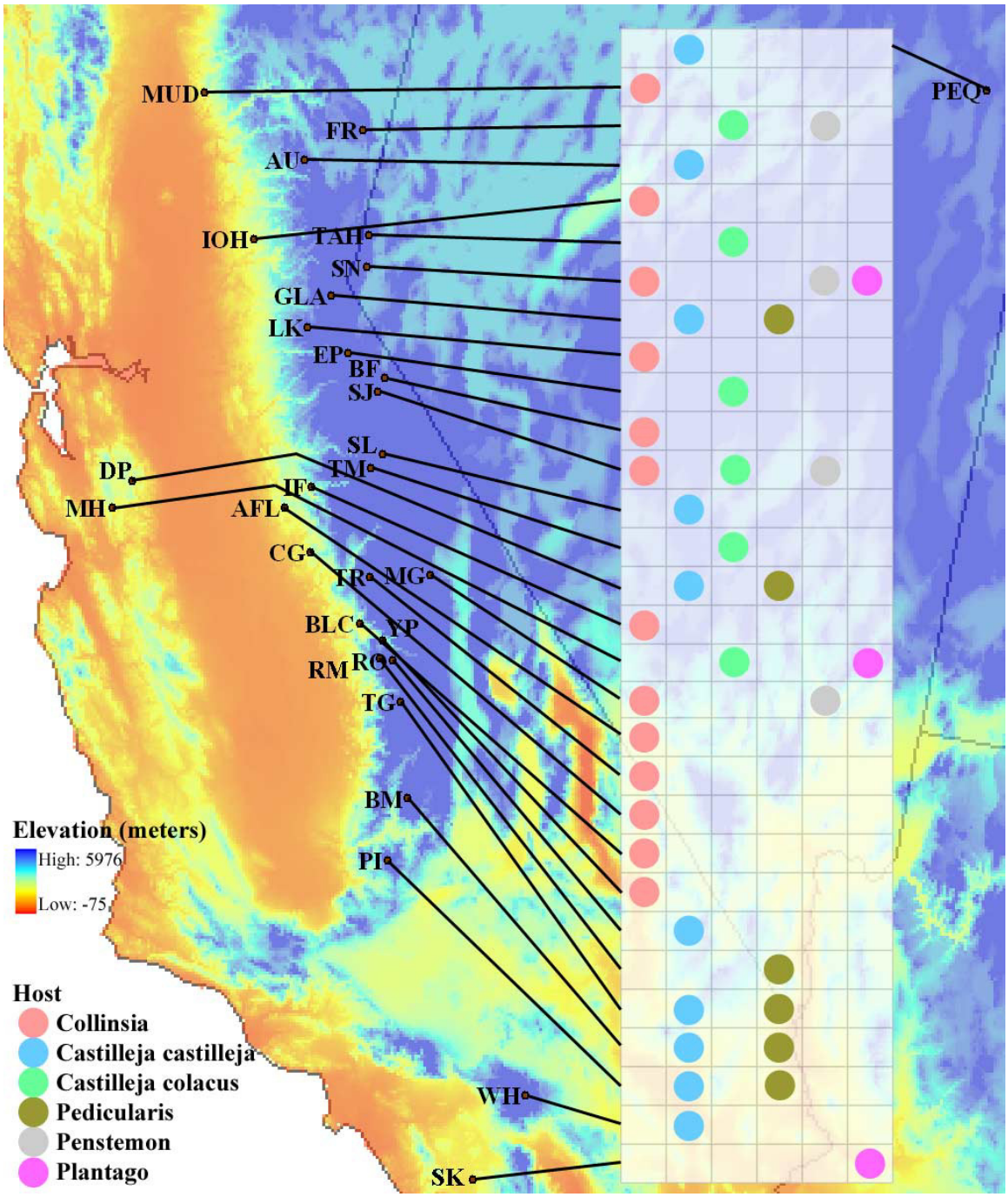


Figure 5.2: NMDS solution for populations at the regional level.

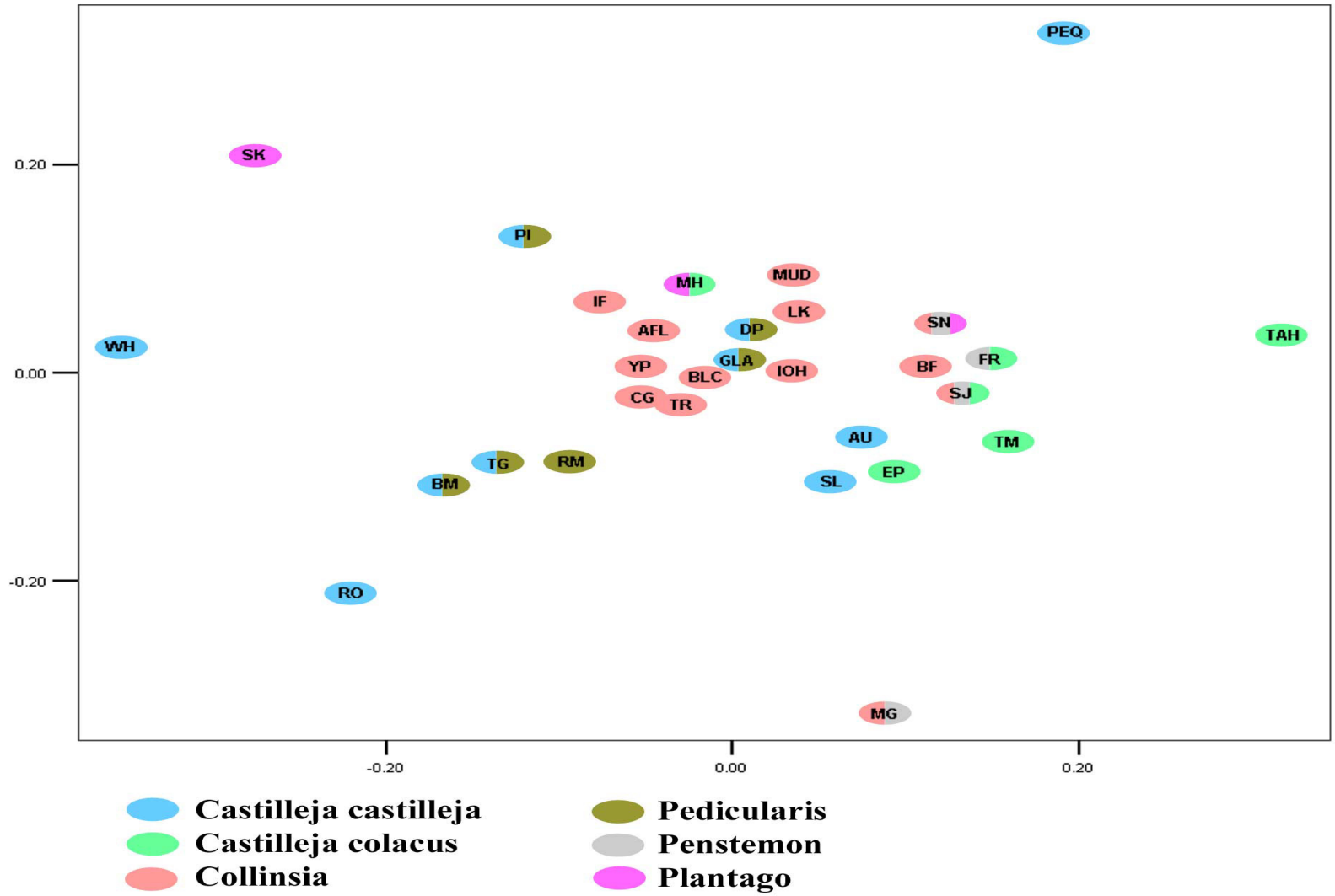


Figure 5.3: UPGMA solution for populations at the regional level.

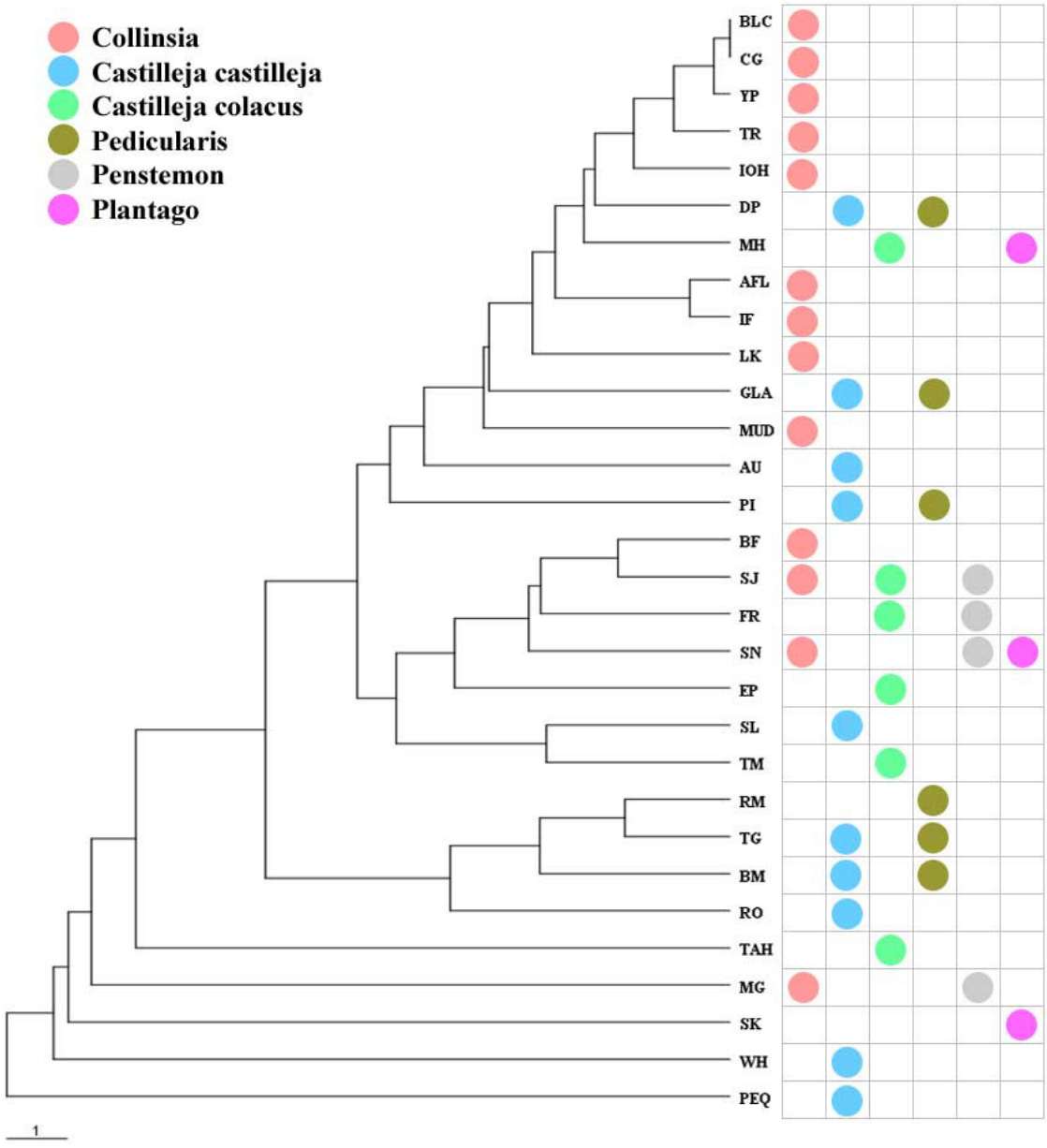


Figure 5.4: IBD plot for populations at the regional level.

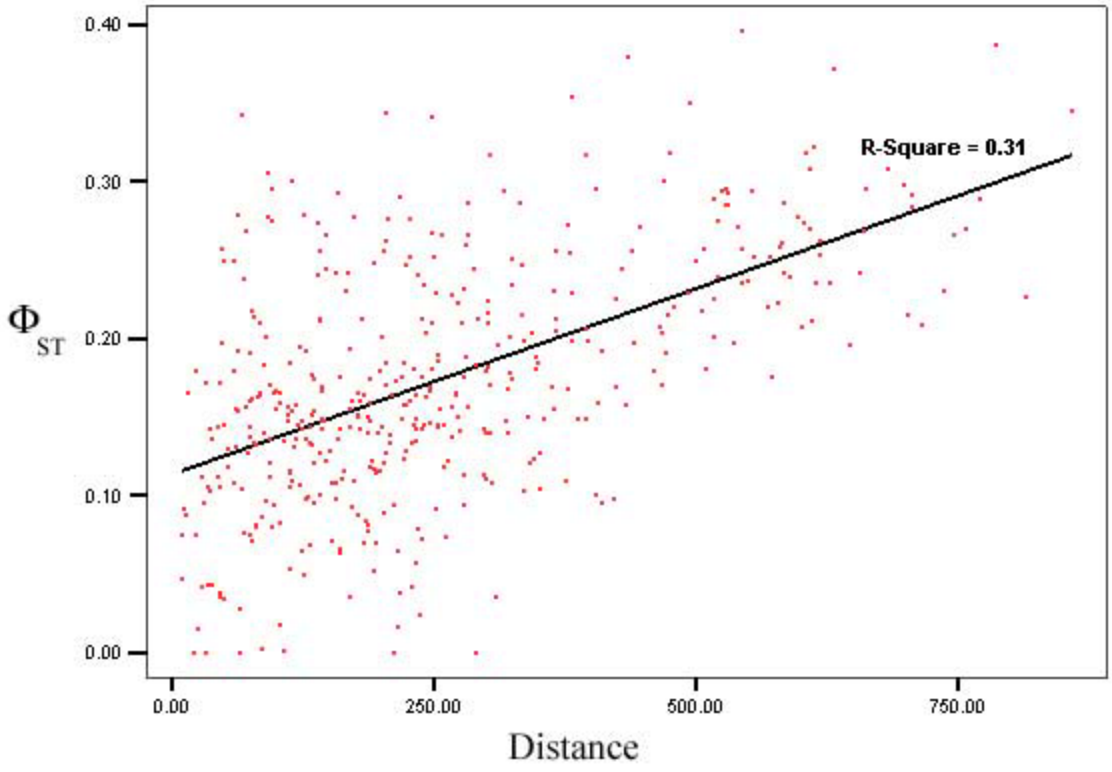
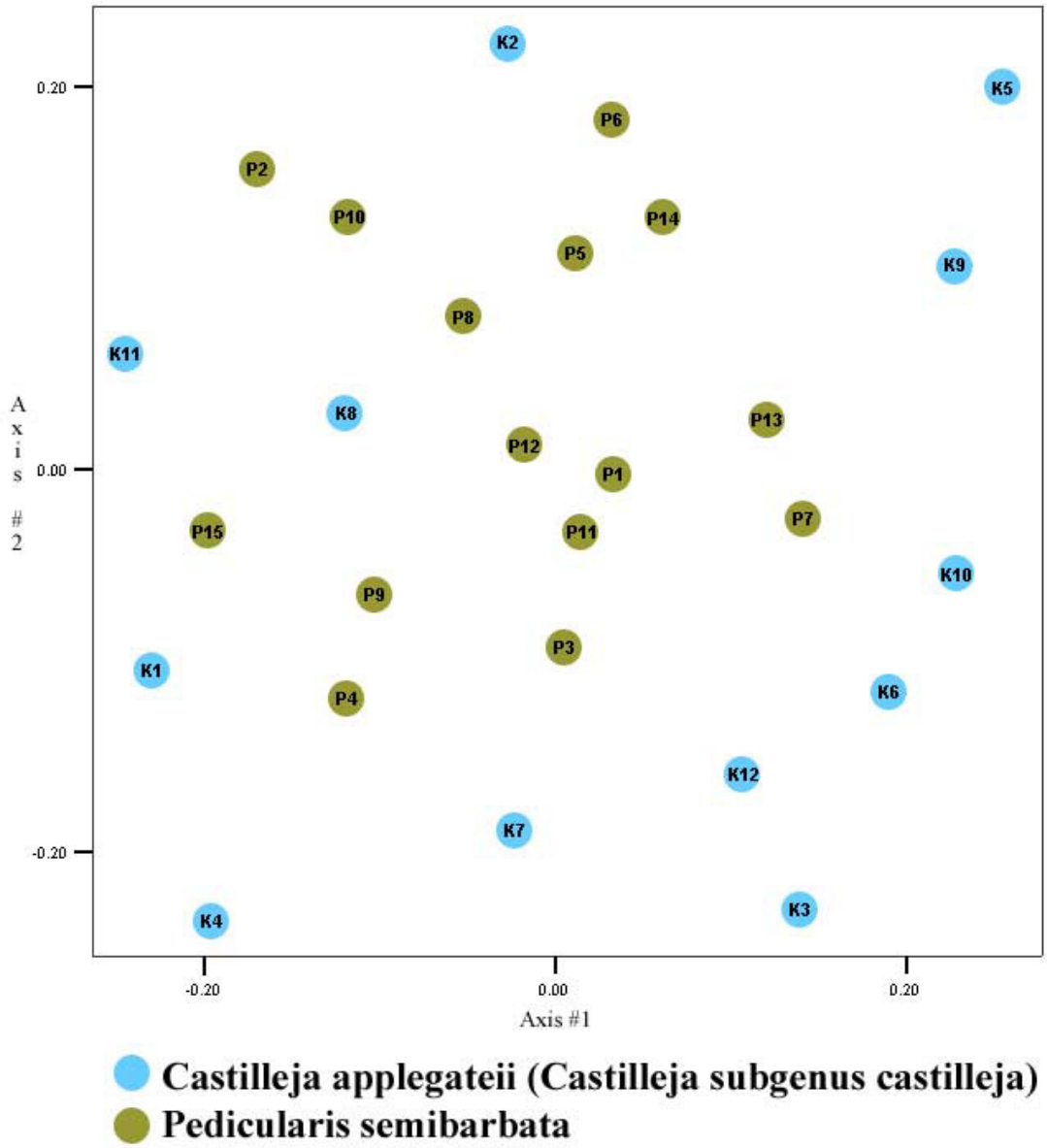


Figure 5.5: NMDS solution for individuals at TJ, a SEKI population.



References

Abrahamson, W. G., M. D. Eubanks, et al. (2001). "Gall flies, inquilines, and goldenrods: A model for host-race formation and sympatric speciation." American Zoologist **41**(4): 928-938.

Anderson, P. R. and J. G. Oakeshott (1984). "Parallel geographical patterns of allozyme variation in two sibling *Drosophila* species." Nature **308**: 729-731.

Baughman, J. F., P. F. Brussard, et al. (1967). "History, selection, drift, and gene flow: complex differentiation in checkerspot butterflies." Canadian Journal of Zoology **68**: 1967-1975.

Baughman, J. F., P. F. Brussard, et al. (1990). "History, selection, drift, and gene flow: complex differentiation in checkerspot butterflies." Canadian Journal of Zoology **68**: 1967-1975.

Begun, D. J. and C. F. Aquadro (1994). "Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation." Genetics **136**: 155-171.

Berlocher, S. H. (1999). "Host race or species? Allozyme characterization or the 'flowering dogwood fly', a member of the *Rhagoletis pomonella* complex." Heredity **83**: 652-662.

Bonnet, E. and Y. Van de Peer (2002). "zt: a software tool for simple and partial Mantel tests." Journal of Statistical software **7**(10): 1-12.

Boughton, D. A. (1999). "Empirical evidence for complex source-sink dynamics with alternative states in a butterfly metapopulation." Ecology **80**(8): 2727-2739.

Boughton, D. A. (2000). "The dispersal system of a butterfly: A test of source-sink theory suggests the intermediate-scale hypothesis." American Naturalist **156**(2): 131-144.

Britten, H. B., P. F. Brussard, et al. (1995). "A test for isolation by distance in central Rocky Mountain populations of Edith's checkerspot butterfly (*Euphydryas editha*)." Heredity **86**: 204-210.

Brussard, P. F., P. Ehrlich, et al. (1985). "Genetic distances and the taxonomy of checkerspot butterflies (Nymphalidae: Nymphalinae)." J. Kansas Entomol. Soc. **58**: 403-412.

Costa, J. and K. Ross (1994). "Hierarchical Genetic-Structure And Gene Flow In Macrogeographic Populations Of The Eastern Tent Caterpillar (*Malacosoma-Americanum*)." Evolution **48**(4): 1158-1167.

Ehrlich, P. (1965). "The population biology of the butterfly *Euphydryas editha.*" Evolution **19**: 327-336.

Excoffier, L. (2001). Analysis of Population Subdivision. Handbook of Statistical Genetics. D. J. Balding, M. Bishop and C. Cannings, John Wiley & Sonds, Ltd**:** 271-307.

Excoffier, L. (2004). Arlequin Frequently Asked Question (FAQ) List, http://lgb.unige.ch/arlequin/software/2.000/doc/faq/faqlist.htm.

Excoffier, L., P. E. Smouse, et al. (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: applicable to human mitochondrial DNA restriction data." Genetics **131**: 479-491.

Feder, J. L. (1998). The apple maggot fly, *Rhagoletis pomonella:* flies in the face of conventional wisdom about speciation? Endless Forms, Species and Speciation. D. J. Howard and S. H. Berlocher. New York, Oxford, Oxford University Press**:** 130-144.

Feder, J. L. and K. E. Filchak (1999). "It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella,* and its implications for fitness trade-offs in phytophagous insects." Entomologia Experimentalis et Applicata **91**: 211-225.

Feder, J. L., T. Hunt, et al. (1993). "The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella.*" Entomologia Experimentalis et Applicata **69**: 117-135.

Feder, J. L., S. Opp, et al. (1994). "Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly." Proc Natl Acad Sci USA **91**: 7990-7994.

Fox, C. (1993). "A quantitative genetic analysis of oviposition preference and larval performance in two hosts in the bruchid beetle *Callosobruchus maculatus*." Evolution **47**: 166-175.

Funk, D. J., K. E. Filchak, et al. (2002). "Herbivorous Insects: Model Systems for the Comparative Study of Speciation Ecology." Genetica **116**: 251-267.

Futuyma, D. J. and G. Moreno (1988). "The evolution of ecological specialization." Annual Review Of Ecology And Systematics **19**: 207-233.

Guiller, A., A. Bellido, et al. (1998). "Genetic Distances and Ordination: The Land Snail *Helix aspersa* in North Africa as a Test Case." Systematic Biology **47**(2): 208-227.

Hanski, I. and M. Singer (2001). "Extinction-Colonization Dynamics and Host-Plant Choice in Butterfly Metapopulations." American Naturalist **158**: 341-353.

Harrison, S. (1989). "Long-distance dispersal and colonization in the bay checkerespot butterfly, *Euphydryas editha bayensis*." Ecology **70**: 1236-1243.

Hartl, D. L. and A. G. Clark (1997). Principles of Population Genetics. Sunderland, MA., Sinauer Associates.

Hawthorne, D. and S. Via (2001). "Genetic linkage of ecological specialization and reproductive isolation in pea aphids." Nature **412**: 904-907.

Hedrick, P. (1999). "Perspective: Highly variable loci and their interpretation in evolution and conservation." Evolution **53**(2): 313-318.

Hintze, J. (2001). NCSS and PASS., Number Cruncher Statistical Systems. Kaysville, Utah. http://www.ncss.com.

Hood, G. M. (2003). PopTools., http://www.cse.csiro.au/poptools.

Jaenike, J. (1990). "Host Specialization In Phytophagous Insects." Annual Review Of Ecology And Systematics **21**: 243-273.

Kawecki, T. J. (1998). "Red queen meets Santa Rosalia:Arms races and the evolution of host specialization in organisms with parasitic lifestyles." American Naturalist **152**: 635-651.

Keyghobadi, N., J. Roland, et al. (1999). "Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus (Papilionidae)*." Molecular Ecology **8**: 1481-1495.

Krebs, C. J. (1989). Ecological methodology. New York, Harper Collins Publishers.

Lessa, E. P. (1990). "Multidimensional analysis of geographical genetic structure." Syst. Zool. **39**(3): 242-252.

Luetgert, J. Range: Distance/Azimuth Program, U.S. Geological Survey; http://quake.wr.usgs.gov/~luetgert/Macintosh_Software/.

Lynch, M. and B. G. Milligan (1994). "Analysis of population genetic structure with RAPD markers." Molecular Ecology **3**: 91-99.

McKechnie, S. W., P. R. Ehrlich, et al. (1975). "Population Genetics of *Euphydryas* Butterflies." Genetics **81**: 571-591.

Mopper, S. and S. Y. Strauss (1998). Genetic Structure and Local Adaptation in Natural Insect Populations. New York, Chapman & Hall.

Nei, M. and W. H. Li (1979). "Mathematical model for studying genetic variation in terms of restriction endonucleases." Proc. Natl. Acad. Sci. USA **76**: 5269-5273.

Ng, D. (1988). "A novel level of interactions in plant-insect systems." Nature **334**: 611-612.

PE-Biosystems Genotyper.

Peterson, M. (1995). "Phenological isolation, gene flow and developmental differences among low and high elevation populations of *Euphilotes enoptes (Lepidoptera: Lycaenidae).*" Evolution **50**: 1990-1999.

Peterson, M. (1996). "Long-distance gene flow in the sedentary butterfly, *Euphilotes enoptes (Lepidoptera: Lycaenidae).*" Evolution **50**(5): 1990-1999.

Peterson, M. and R. Denno (1998). "The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects." American Naturalist **152**(3): 428-446.

Peterson, M., R. Denno, et al. (2001). "Apparent widespread gene flow in the predominantly flightless planthopper *Tumidagena minuta.*" Ecological Entomology **26**(6): 629-637.

Radtkey, R. and M. C. Singer (1995). "Repeated reversals of host-preference evolution in a specialist insect herbivore." Evolution **49**: 351-359.

Rausher, M. D. (1982). "Population differentiation in *Euphydryas editha*: Larval adaptation to different hosts." Evolution **36**: 581-590.

Saccheri, I., I. Wilson, et al. (1999). "Inbreeding of bottlenecked butterfly populations: Estimation using the likelihood of changes in marker allele frequencies." Genetics **151**: 1053-1063.

Schneider, S., D. Roessli, et al. (2000). Arlequin: A software for population genetics data analysis., Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Singer, M. C. (1983). "Determinants of multiple host use by a phytophagous insect population." Evolution **37**: 389-403.

Singer, M. C. (1994). Behavioral constraints on the evolutionary expansion of insect diet: a case history from checkerspot butterflies. Behavioral Mechanisms in Evolutionary Ecology. L. Real, University of Chicago Press**:** 279-296.

Singer, M. C. (2004). Measurement, Correlates, and Importance of Oviposition Preference in the Life of Cherkerspots. On the Wings of Checkerspots. P. Ehrlich and I. Hanski. Oxford, NY, Oxford University Press.

Singer, M. C., D. Ng, et al. (1988). "Heritability of oviposition preference and its relationship to offspring performance within a single insect population." Evolution **42**: 977-985.

Singer, M. C. and C. Parmesan (1993). "Sources of variation in patterns of plant-insect association." Nature **361**: 251-253.

Singer, M. C. and C. D. Thomas (1996). "Evolutionary responses of a butterfly metapopulation to human and climate-caused environmental variation." American Naturalist **148**: S9-S39.

Singer, M. C., C. D. Thomas, et al. (1993). "Rapid human-induced evolution of insect diet." Nature **366**: 681-683.

Slatkin, M. (1987). "Gene flow and the geographic structure of natural populations." Science **236**: 787-792.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts., Sinauer Associates.

Thomas, C. D. and M. C. Singer (1998). Scale-dependent evolution of specialization in a checkerspot butterfly. Genetic Structure and local adaptation in natural insect populations. S. a. S. Y. S. Mopper, Chapman & Hall**:** 343-374.

Thomas, C. D., M. C. Singer, et al. (1996). "Catastrophic extinction of population sources in a butterfly metapopulation." American Naturalist **148**: 957-975.

Thomas, Y., M. Bethenod, et al. (2003). "Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis hubner*. I. Sex pheromone, moth emergence timing, and parasitism." Evolution **57**(2): 261-273.

Vekemans, X. (2002). AFLP-SURV, Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via, S. (1986). "Genetic covariance between oviposition preference and larval performance in an insect herbivore." Evolution **40**: 778-785.

Via, S. (1999). "Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice." Evolution **53**(5): 1446-1457.

Wahlberg, N. and M. Zimmermann (2000). "Pattern of Phylogenetic Relationships among Members of the Tribe *Melitaeini (Lepidoptera: Nymphalidae)* Inferred from Mitochondrial DNA Sequences." Cladistics **16**: 347-363.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**: 1358 - 1370.

Wood, T. K. and M. C. Keese (1990). "Host plant induced assortative mating in *Enchenopa* treehoppers." Evolution **44**: 619-628.

Wright, S. (1943). "Isolation by distance." Genetics **28**: 139-156.

Wright, S. (1978). Evolution and the Genetics of Populations. Chicago, University of Chicago Press.

Wu, J., S. Huang, et al. (2001). "Allozyme variation and the genetic structure of populations of *Trochodendron aralioides*, a monotypic and narrow geographic genus." Journal Of Plant Research **114**: 45-57.

Zhivotovsky, L. A. (1999). "Estimating population structure in diploids with multilocus dominant DNA markers." Molecular Ecology **8**: 907-913.

Zimmermann, M., N. Wahlberg, et al. (2000). "Phylogeny of Euphydryas Checkerspot Butterflies (*Lepidoptera: Nymphalidae*) Based on Mitochondrial DNA Sequence Data." Ann. Entomol. Soc. Am. **93**(3): 347-355.

CONSOLIDATED REFERENCES

Abrahamson, W. G., M. D. Eubanks, et al. (2001). "Gall flies, inquilines, and goldenrods: A model for host-race formation and sympatric speciation." American Zoologist **41**(4): 928-938.

Anderson, P. R. and J. G. Oakeshott (1984). "Parallel geographical patterns of allozyme variation in two sibling *Drosophila* species." Nature **308**: 729-731.

Anthes, N., T. Fartmann, et al. (2003). "Combining larval habitat quality and metapopulation structure – the key for successful management of pre-alpine *Euphydryas aurinia* colonies." Journal of Insect Conservation **7**(3): 175-185.

Asher, J., M. Warren, et al. (2001). The Millennium Atlas of Butterflies in Britain and Ireland, Oxford University Press.

Baughman, J. F., P. F. Brussard, et al. (1967). "History, selection, drift, and gene flow: complex differentiation in checkerspot butterflies." Canadian Journal of Zoology **68**: 1967-1975.

Baughman, J. F., P. F. Brussard, et al. (1990). "History, selection, drift, and gene flow: complex differentiation in checkerspot butterflies." Canadian Journal of Zoology **68**: 1967-1975.

Begun, D. J. and C. F. Aquadro (1994). "Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation." Genetics **136**: 155-171.

Berlocher, S. H. (1999). "Host race or species? Allozyme characterization or the 'flowering dogwood fly', a member of the *Rhagoletis pomonella* complex." Heredity **83**: 652-662.

Berlocher, S. H. and J. L. Feder (2002). "Sympatric Speciation in Phytophagous Insects: Moving Beyond Controversy?" Annual Review of Entomology **47**: 773-815.

Bonnet, E. and Y. Van de Peer (2002). "zt: a software tool for simple and partial Mantel tests." Journal of Statistical software **7**(10): 1-12.

Boughton, D. A. (1999). "Empirical evidence for complex source-sink dynamics with alternative states in a butterfly metapopulation." Ecology **80**(8): 2727-2739.

Boughton, D. A. (2000). "The dispersal system of a butterfly: A test of source-sink theory suggests the intermediate-scale hypothesis." American Naturalist **156**(2): 131-144.

Britten, H. B., P. F. Brussard, et al. (1995). "A test for isolation by distance in central Rocky Mountain populations of Edith's checkerspot butterfly (*Euphydryas editha*)." Heredity **86**: 204-210.

Brussard, P. F., P. Ehrlich, et al. (1985). "Genetic distances and the taxonomy of checkerspot butterflies (Nymphalidae: Nymphalinae)." J. Kansas Entomol. Soc. **58**: 403-412.

Bush, G. L. (1969). "Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis (Diptera: Tephritidae)*." Evolution **23**: 237-251.

Bush, G. L. and J. J. Smith (1998). "The genetics and ecology of sympatric speciation: a case study." Res. Popul. Ecol. **40**: 175-187.

Caillaud, C. M. and S. Via (2000). "Specialised feeding behaviour influences both ecological specialisation and assortative mating in sympatric host races of pea aphids." American Naturalist **156**: 606-621.

Collischonn, W. and J. V. Pilar (2000). "A direction dependent least-cost path algorithm for roads and canals." International Journal of Geographic Information Systems **14**(4): 397 - 406.

Costa, J. and K. Ross (1994). "Hierarchical Genetic-Structure And Gene Flow In Macrogeographic Populations Of The Eastern Tent Caterpillar (*Malacosoma-Americanum*)." Evolution **48**(4): 1158-1167.

Craig, T. P., J. D. Horner, et al. (1997). "Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation." Evolution **51**(5): 1552-1560.

Craig, T. P., J. D. Horner, et al. (2001). "Genetics, experience, and host-plant preference in *Eurosta solidaginis*: implications for host shifts and speciation." Evolution **55**(4): 773-782.

Craig, T. P., J. K. Itami, et al. (2000). "The influence of host plant variation and intraspecific competition on oviposition preference and offspring performance in the host races of *Eurosta solidaginis*." Ecological Entomology **25**: 1-12.

Del Campo, M. L., S. Via, et al. (2003). "Recognition of host-specific chemical stimulants in two sympatric host races of the pea aphid *Acyrthosiphon pisum.*" Ecological Entomology **28**(4): 405-412.

Douglas, D. H. (1994). "Least-cost Path in GIS Using an Accumulated Cost Surface and Slopelines." Cartographica **31**: 37-51.

Dres, M. and J. Mallet (2001). "Host races in plant-feeding insects and their importance in sympatric speciation." Phil. Trans. R. Soc. Lond. B **357**: 471-492.

Ehrlich, P. (1965). "The population biology of the butterfly *Euphydryas editha.*" Evolution **19**: 327-336.

ESRI (2002). ArcGIS 8.1, ESRI Inc.

ESRI (2002). ArcGIS Spatial Analyst Functional Reference, ESRI Inc.

Excoffier, L. (2001). Analysis of Population Subdivision. Handbook of Statistical Genetics. D. J. Balding, M. Bishop and C. Cannings, John Wiley & Sonds, Ltd**:** 271-307.

Excoffier, L. (2004). Arlequin Frequently Asked Question (FAQ) List, http://lgb.unige.ch/arlequin/software/2.000/doc/faq/faqlist.htm.

Excoffier, L., P. E. Smouse, et al. (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: applicable to human mitochondrial DNA restriction data." Genetics **131**: 479-491.

Fahrig, L. (2001). "How much habitat is enough?" Biological Conservation **100**(1): 65-74.

Fahrig, L. and G. Merriam (1994). "Conservation of Fragmented Populations." Conservation Biology **8**(1): 50-59.

Feder, J. L. (1998). The apple maggot fly, *Rhagoletis pomonella:* flies in the face of conventional wisdom about speciation? Endless Forms, Species and Speciation. D. J. Howard and S. H. Berlocher. New York, Oxford, Oxford University Press**:** 130-144.

Feder, J. L. and K. E. Filchak (1999). "It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella,* and its implications for fitness trade-offs in phytophagous insects." Entomologia Experimentalis et Applicata **91**: 211-225.

Feder, J. L., S. Opp, et al. (1994). "Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly." Proc Natl Acad Sci USA **91**: 7990-7994.

Feder, J. L., T. Hunt, et al. (1993). "The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella.*" Entomologia Experimentalis et Applicata **69**: 117-135.

Feehan, J. and G. O'Donovan (1996). The Bogs of Ireland. Dublin, Ireland, University College Dublin.

Filchak, K. E., J. B. Roethele, et al. (2000). "Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*." Nature **407**: 739-742.

Fox, C. (1993). "A quantitative genetic analysis of oviposition preference and larval performance in two hosts in the bruchid beetle *Callosobruchus maculatus*." Evolution **47**: 166-175.

Funk, D. J., K. E. Filchak, et al. (2002). "Herbivorous Insects: Model Systems for the Comparative Study of Speciation Ecology." Genetica **116**: 251-267.

Futuyma, D. J. and G. Moreno (1988). "The evolution of ecological specialization." Annual Review Of Ecology And Systematics **19**: 207-233.

Guiller, A., A. Bellido, et al. (1998). "Genetic Distances and Ordination: The Land Snail *Helix aspersa* in North Africa as a Test Case." Systematic Biology **47**(2): 208-227.

Hanski, I. and M. Singer (2001). "Extinction-Colonization Dynamics and Host-Plant Choice in Butterfly Metapopulations." American Naturalist **158**: 341-353.

Hanski, I. and O. Gaggiotti (2004). Ecology, Genetics and Evolution of Metapopulations., Academic Press.

Harrison, S. (1989). "Long-distance dispersal and colonization in the bay checkerespot butterfly, *Euphydryas editha bayensis*." Ecology **70**: 1236-1243.

Hartl, D. L. and A. G. Clark (1997). Principles of Population Genetics. Sunderland, MA., Sinauer Associates.

Hawthorne, D. and S. Via (2001). "Genetic linkage of ecological specialization and reproductive isolation in pea aphids." Nature **412**: 904-907.

Hedrick, P. (1999). "Perspective: Highly variable loci and their interpretation in evolution and conservation." Evolution **53**(2): 313-318.

Henein, K. and G. Merriam (1990). "The elements of connectivity where corridor quality is variable." Landscape Ecology **4**: 147-170.

Hintze, J. (2001). NCSS and PASS., Number Cruncher Statistical Systems. Kaysville, Utah. http://www.ncss.com.

Hood, G. M. (2003). PopTools., http://www.cse.csiro.au/poptools.

Itami, J. K., T. P. Craig, et al. (1997). Factors affecting gene flow between the host races of *Eurosta solidaginis*. Genetic structure and local adaptation in natural insect populations: Effects of ecology, life history, and behavior. S. Mopper and S. Y. Strauss. New York, Chapman & Hall**:** 375-407.

Jaenike, J. (1990). "Host Specialization In Phytophagous Insects." Annual Review Of Ecology And Systematics **21**: 243-273.

Jones, C. J., K. J. Edwards, et al. (1997). "Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories." Molecular Breeding **3**(5): 381-390.

Joyce, D. A. and A. S. Pullin (2001). "Phylogeography of the Marsh Fritillary *Euphydryas aurinia (Lepidoptera: Nymphalidae)* in the UK." Biological Journal of the Linnean Society **72**: 129-141.

Kawecki, T. J. (1998). "Red queen meets Santa Rosalia:Arms races and the evolution of host specialization in organisms with parasitic lifestyles." American Naturalist **152**: 635-651.

Keyghobadi, N., J. Roland, et al. (1999). "Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus (Papilionidae)*." Molecular Ecology **8**: 1481-1495.

Krebs, C. J. (1989). Ecological methodology. New York, Harper Collins Publishers.

Kuussaari, M., M. Singer, et al. (2000). "Local specialization and landscape-level influence on host use in an herbivorous insect." Ecology **81**(8): 2177-2187.

Lessa, E. P. (1990). "Multidimensional analysis of geographical genetic structure." Syst. Zool. **39**(3): 242-252.

Levins, R. (1969). "Some demographic and genetic consequences of environmental heterogeneity for biological control." Bulletin of the Entomological Society of America **15**: 237-240.

Luetgert, J. Range: Distance/Azimuth Program, U.S. Geological Survey; http://quake.wr.usgs.gov/~luetgert/Macintosh_Software/.

Lynch, M. and B. G. Milligan (1994). "Analysis of population genetic structure with RAPD markers." Molecular Ecology **3**: 91-99.

MacArthur, R. and E. Wilson (1963). "An equilibrium theory of insular zoogeography." Evolution **17**: 373-387.

MacArthur, R. and E. Wilson (1967). The Theory of Island Biogeography. Princeton, New Jersey., Princeton University Press.

Mazel, R. (1986). "Structure et evolution du peuplement d'*Euphydryas aurinia* Rottemburg (Lepidoptera) dans le sud-ouest europeen." Vie Milieu **36**: 205-225.

McKechnie, S. W., P. R. Ehrlich, et al. (1975). "Population Genetics of *Euphydryas* Butterflies." Genetics **81**: 571-591.

Michels, E., K. Cottenie, et al. (2001). "Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: a plea for using GIS modelling of the effective geographical distance." Molecular Ecology **10**: 1929-1938.

Moilanen, A. and I. Hanski (1998). "Metapopulation Dynamics: Effects of Habitat Quality and Landscape Structure." Ecology **79**(7): 2503-2515.

Mopper, S. and S. Y. Strauss (1998). Genetic Structure and Local Adaptation in Natural Insect Populations. New York, Chapman & Hall.

Mueller, U. G. and L. Wolfenbarger (1999). "AFLP genotyping and fingerprinting." Trends in Ecology and Evolution **14**: 389-393.

Nei, M. and W. H. Li (1979). "Mathematical model for studying genetic variation in terms of restriction endonucleases." Proc. Natl. Acad. Sci. USA **76**: 5269-5273.

Ng, D. (1988). "A novel level of interactions in plant-insect systems." Nature **334**: 611-612.

Norberg, U., K. Enfjall, et al. (2002). "Habitat exploration in butterflies – an outdoor cage experiment." Evolutionary Ecology **16**: 1-14.

Parmesan, C., M. C. Singer, et al. (1995). "Absence of adaptive learning from the oviposition foraging behaviour of a checkerspot butterfly." Animal Behaviour **50**: 161-175.

PE-Biosystems Genotyper.

Peterson, M. (1995). "Phenological isolation, gene flow and developmental differences among low and high elevation populations of *Euphilotes enoptes (Lepidoptera: Lycaenidae).*" Evolution **50**: 1990-1999.

Peterson, M. (1996). "Long-distance gene flow in the sedentary butterfly, *Euphilotes enoptes (Lepidoptera: Lycaenidae).*" Evolution **50**(5): 1990-1999.

Peterson, M. (1997). "Host plant phenology and butterfly dispersal: Causes and consequences of uphill movement." Ecology **78**(1): 167-180.

Peterson, M. A. and R. F. Denno (1998). Life-History Strategies and the Genetic Structure of Phytophagous Insect Populations. Genetic Structure and local adaptation in natural insect populations. S. Mopper and S. SY, Chapman & Hall.

Peterson, M. and R. Denno (1998). "The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects." American Naturalist **152**(3): 428-446.

Peterson, M., R. Denno, et al. (2001). "Apparent widespread gene flow in the predominantly flightless planthopper *Tumidagena minuta.*" Ecological Entomology **26**(6): 629-637.

Radtkey, R. and M. C. Singer (1995). "Repeated reversals of host-preference evolution in a specialist insect herbivore." Evolution **49**: 351-359.

Rausher, M. D. (1982). "Population differentiation in *Euphydryas editha*: Larval adaptation to different hosts." Evolution **36**: 581-590.

Ricketts, T. H. (2001). "The Matrix Matters: Effective Isolation in Fragmented Landscapes." American Naturalist **158**(1): 87-99.

Roland, J., N. Keyghobadi, et al. (2000). "Alpine Parnassius butterfly dispersal: Effects of landscape and population size." Ecology **81**(6): 1642-1653.

Saccheri, I., I. Wilson, et al. (1999). "Inbreeding of bottlenecked butterfly populations: Estimation using the likelihood of changes in marker allele frequencies." Genetics **151**: 1053-1063.

Schneider, S., D. Roessli, et al. (2000). Arlequin: A software for population genetics data analysis., Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Simberloff, D. S. (1974). "Equilibrium theory of island biogeography and ecology." Annual Review of Ecology and Systematics **5**: 161-182.

Singer, M. C. (1983). "Determinants of multiple host use by a phytophagous insect population." Evolution **37**: 389-403.

Singer, M. C. (1994). Behavioral constraints on the evolutionary expansion of insect diet: a case history from checkerspot butterflies. Behavioral Mechanisms in Evolutionary Ecology. L. Real, University of Chicago Press**:** 279-296.

Singer, M. C. (2004). Measurement, Correlates, and Importance of Oviposition Preference in the Life of Cherkerspots. On the Wings of Checkerspots. P. Ehrlich and I. Hanski. Oxford, NY, Oxford University Press.

Singer, M. C. and C. D. Thomas (1996). "Evolutionary responses of a butterfly metapopulation to human and climate-caused environmental variation." American Naturalist **148**: S9-S39.

Singer, M. C. and C. Parmesan (1993). "Sources of variation in patterns of plant-insect association." Nature **361**: 251-253.

Singer, M. C., C. D. Thomas, et al. (1993). "Rapid human-induced evolution of insect diet." Nature **366**: 681-683.

Singer, M. C., C. Stefanescu, et al. (2002). "When random sampling does not work: standard design falsely indicates maladaptive host preferences in a butterfly." Ecology Letters, **5**: 1-6.

Singer, M. C., D. Ng, et al. (1988). "Heritability of oviposition preference and its relationship to offspring performance within a single insect population." Evolution **42**: 977-985.

Slatkin, M. (1987). "Gene flow and the geographic structure of natural populations." Science **236**: 787-792.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts., Sinauer Associates.

Thomas, C. D. and M. C. Singer (1998). Scale-dependent evolution of specialization in a checkerspot butterfly. Genetic Structure and local adaptation in natural insect populations. S. a. S. Y. S. Mopper, Chapman & Hall**:** 343-374.

Thomas, C. D., M. C. Singer, et al. (1996). "Catastrophic extinction of population sources in a butterfly metapopulation." American Naturalist **148**: 957-975.

Thomas, Y., M. Bethenod, et al. (2003). "Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis hubner*. I. Sex pheromone, moth emergence timing, and parasitism." Evolution **57**(2): 261-273.

Turner, M., R. Gardner, et al. (2001). Landscape Ecology in Theory and Practice: Pattern and Process. New York, New York., Springer-Verlag.

Vekemans, X. (2002). AFLP-SURV, Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via, S. (1986). "Genetic covariance between oviposition preference and larval performance in an insect herbivore." Evolution **40**: 778-785.

Via, S. (1991). "Specialised host plant performance is not altered by experience." Ecology **72**: 1420-1427.

Via, S. (1991). "The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones." Evolution **45**: 827-852.

Via, S. (1999). "Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice." Evolution **53**(5): 1446-1457.

Via, S., A. C. Bouck, et al. (2000). "Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environment." Evolution **54**(5): 1626-1637.

Wahlberg, N. and M. Zimmermann (2000). "Pattern of Phylogenetic Relationships among Members of the Tribe *Melitaeini (Lepidoptera: Nymphalidae)* Inferred from Mitochondrial DNA Sequences." Cladistics **16**: 347-363.

Wahlberg, N., T. Klemetti, et al. (2002). "Dynamic populations in a dynamic landscape: The metapopulation structure of the marsh fritillary butterfly." Ecography **25**(2): 224-232.

Wahlberg, N., T. Klemetti, et al. (2002). "Metapopulation structure and movements in five species of checkerspot butterflies." Oecologia **130**: 33-43.

Watt, W. B. (1969). "Adaptive significance of pigment polymorphisms in *Colias* butterflies. II. Thermoregulation and photoperiodically controlled melanin variation in *Colias eurytheme*." Proceedings of the National Academy of Sciences of the United States of America **63**: 767-774.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**: 1358 - 1370.

Weiss, S. B. and W. A.D. (1998). "Landscape-Level Phenology of a Threatened Butterfly: A GIS-Based Modeling Approach." Ecosystems **1**: 299-309.

Wood, T. K. and M. C. Keese (1990). "Host plant induced assortative mating in *Enchenopa* treehoppers." Evolution **44**: 619-628.

Wright, S. (1943). "Isolation by distance." Genetics **28**: 139-156.

Wright, S. (1978). Evolution and the Genetics of Populations. Chicago, University of Chicago Press.

Wu, J., S. Huang, et al. (2001). "Allozyme variation and the genetic structure of populations of *Trochodendron aralioides*, a monotypic and narrow geographic genus." Journal Of Plant Research **114**: 45-57.

Zhivotovsky, L. A. (1999). "Estimating population structure in diploids with multilocus dominant DNA markers." Molecular Ecology **8**: 907-913.

Zimmermann, M., N. Wahlberg, et al. (2000). "Phylogeny of Euphydryas Checkerspot Butterflies (*Lepidoptera: Nymphalidae*) Based on Mitochondrial DNA Sequence Data." Ann. Entomol. Soc. Am. **93**(3): 347-355.

VITA

Peck Sheng Wee was born in Singapore in April 15, 1968, the son of Wee Chew Beng and Yeo Guat Eng. He was awarded a full scholarship for his studies at the National Universty of Singapore in 1989 by Neptune Orient Line, Singapore's national shipping line. After receiving his Bachelor of Science (Information Systems and Computer Science) in 1992, he worked for Andersen Consulting, a global information technology and management consulting firm. After two years, he was sponsored by Andersen Consulting for his Master of Science (Computer Science) at Northwestern University, Evanston, Illinois. After graduation, he returned to work with Andersen Consulting for an additional two years. In 1999, he entered the Graduate School of The University of Texas at Austin.

Permanent Address: Blk 503, Jelapang Rd, #08-362, Singapore 670503.

This dissertation was typed by the author.